

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2005, 01:08:33 ; Search time 759 Seconds
(without alignments)
8992.812 Million cell updates/sec

Title: US-10-829-118-1
Perfect score: 980
Sequence: 1 ctccgcggcgacacgcctg.....ctgcaccacggaagtctga 980

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 8443130 seqs, 3482420727 residues

Total number of hits satisfying chosen parameters: 16986260

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US05_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
- 18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
- 19: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq:*
- 20: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq:*
- 21: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq:*
- 22: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 23: /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq:*
- 24: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
- 25: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 26: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	980	100.0	980	9	US-09-891-711-1
2	980	100.0	980	19	US-10-829-118-1
3	980	100.0	2469	9	US-09-891-711-2
4	980	100.0	2469	19	US-10-829-118-2
c 5	124.4	12.7	4338	9	US-09-891-711-3
c 6	124.4	12.7	4338	17	US-10-172-118-953
c 7	124.4	12.7	4338	18	US-10-342-887-953

c 8	124.4	12.7	4338	19	US-10-829-118-3	Sequence 3, Appli
c 9	124.4	12.7	4338	21	US-10-956-157-1116	Sequence 1116, Ap
c 10	56.4	5.8	1523	19	US-10-767-701-9786	Sequence 9786, Ap
c 11	53	5.4	658	19	US-10-767-701-4204	Sequence 4204, Ap
c 12	52.6	5.4	4667	20	US-10-723-860-5759	Sequence 5759, Ap
c 13	51.6	5.3	598	20	US-10-425-115-2989	Sequence 2989, Ap
c 14	51.6	5.3	763	19	US-10-437-963-49370	Sequence 49370, A
c 15	51.4	5.2	606	18	US-10-424-599-95310	Sequence 95310, A
c 16	50.8	5.2	2307	10	US-09-893-519A-87	Sequence 87, Appli
c 17	49.8	5.1	849	20	US-10-363-345A-16551	Sequence 16551, A
c 18	49.8	5.1	849	20	US-10-363-345A-16552	Sequence 16552, A
c 19	49.8	5.1	849	21	US-10-363-483A-16551	Sequence 16551, A
c 20	49.8	5.1	849	21	US-10-363-483A-16552	Sequence 16552, A
c 21	48.8	5.0	771	14	US-10-010-920-4	Sequence 4, Appli
c 22	48.8	5.0	771	14	US-10-010-920-5	Sequence 5, Appli
c 23	48.8	5.0	771	14	US-10-008-721-4	Sequence 4, Appli
c 24	48.8	5.0	771	14	US-10-008-721-5	Sequence 5, Appli
c 25	48.8	5.0	955	14	US-10-010-920-3	Sequence 3, Appli
c 26	48.8	5.0	955	14	US-10-008-721-3	Sequence 3, Appli
c 27	48.8	5.0	40668	19	US-10-450-826-57	Sequence 57, Appli
c 28	48.6	5.0	809	18	US-10-424-599-26988	Sequence 26988, A
c 29	48.4	4.9	4225	16	US-10-141-541-3	Sequence 3, Appli
c 30	48.2	4.9	1187	16	US-10-029-386-25125	Sequence 25125, A
c 31	48	4.9	1092	19	US-10-767-701-9739	Sequence 9739, Ap
c 32	48	4.9	1215	19	US-10-767-701-8767	Sequence 8767, Ap
c 33	48	4.9	38847	13	US-10-087-192-868	Sequence 868, App
c 34	48	4.9	150275	22	US-10-981-277-55	Sequence 55, Appli
c 35	47.8	4.9	951	19	US-10-437-963-63083	Sequence 63083, A
c 36	47.6	4.9	1041	19	US-10-437-963-8597	Sequence 8597, Ap
c 37	47.4	4.8	154746	10	US-09-827-688-8	Sequence 8, Appli
c 38	47.4	4.8	154746	10	US-09-827-688-8	Sequence 8, Appli
c 39	47.2	4.8	825	19	US-10-437-963-60378	Sequence 60378, A
c 40	47	4.8	876	20	US-10-425-115-55664	Sequence 55664, A
c 41	47	4.8	876	20	US-10-425-115-55664	Sequence 55664, A
c 42	46.8	4.8	1211	19	US-10-437-963-27585	Sequence 27585, A
c 43	46.8	4.8	2571	14	US-10-277-032-1	Sequence 1, Appli
c 44	46.8	4.8	2571	18	US-10-681-223-1	Sequence 1, Appli
c 45	46.8	4.8	2856	16	US-10-305-810-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-09-891-711-1
; Sequence 1, Application US/09891711
; Patent No. US20020082404A1
; GENERAL INFORMATION:
; APPLICANT: Pharmacia AB
; TITLE OF INVENTION: Promoter Sequences
; FILE REFERENCE: 00130
; CURRENT APPLICATION NUMBER: US/09/891,711
; CURRENT FILING DATE: 2001-06-26
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 980
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-891-711-1

Query Match 100.0%; Score 980; DB 9; Length 980;
Best Local Similarity 100.0%; Pred. No. 1.7e-288;
Matches 980; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	CTCCGCGCGGACACGCCTGGGCACCTCCATTCCGGGGCTGTTTACTCCCAACTCTCCGCGAG 60
Db	1	CTCCGCGCGGACACGCCTGGGCACCTCCATTCCGGGGCTGTTTACTCCCAACTCTCCGCGAG 60
Qy	61	ACTGGGCGCGCGGCGGCGGAGGCCACAGCTGGGAGCCTCAGCTCCGCGCCGACCCAGCGT 120
Db	61	ACTGGGCGCGCGGCGGCGGAGGCCACAGCTGGGAGCCTCAGCTCCGCGCCGACCCAGCGT 120

```

QY 121 GCCCTGTCTGTCCTCCGCGCTCCCGGGCTTGGTGGCGCTCTGACGCGCTGGGAGCGG 180
DB 121 GCCCTGTCTGTCCTCCGCGCTCCCGGGCTTGGTGGCGCTCTGACGCGCTGGGAGCGG 180
QY 181 GACACGCGGAGGATGGAAGAGTCTCGGACATTTGGCGGGCGGGCGGGTGG 240
DB 181 GACACGCGGAGGATGGAAGAGTCTCGGACATTTGGCGGGCGGGCGGGTGG 240
QY 241 CAGGTTGAAAGCGGAGGCGTGGCCAGCGAGCTGCCAGGCGGAGAACCGCTGGGG 300
DB 241 CAGGTTGAAAGCGGAGGCGTGGCCAGCGAGCTGCCAGGCGGAGAACCGCTGGGG 300
QY 301 AACCTTTGGTCCGCTCTGCGGCTCGCTTAGATCCCGAAAAAGAGCACGGGCGGAAA 360
DB 301 AACCTTTGGTCCGCTCTGCGGCTCGCTTAGATCCCGAAAAAGAGCACGGGCGGAAA 360
QY 361 GCGGCGAGCTGGGCGAGGATCTAGAAAGACTGCTGGCGGAGGCTCCCTGCCCGCGG 420
DB 361 GCGGCGAGCTGGGCGAGGATCTAGAAAGACTGCTGGCGGAGGCTCCCTGCCCGCGG 420
QY 421 GCCTGTCTCATGACTCGTGGAGCTCGCTTCCCGCGGAGCCCTTCTCTGCAAGGGTC 480
DB 421 GCCTGTCTCATGACTCGTGGAGCTCGCTTCCCGCGGAGCCCTTCTCTGCAAGGGTC 480
QY 481 CACGTCCAGGACCGGCGGCTCGGACACCCACCCCGCGGCGGACCTGCTGCGGTGG 540
DB 481 CACGTCCAGGACCGGCGGCTCGGACACCCACCCCGCGGCGGACCTGCTGCGGTGG 540
QY 541 CCTTTAAACCCCGGCGTGGTAAAGTGGCGAAGTGTCCGCTCCGGAACACCGGAAA 600
DB 541 CCTTTAAACCCCGGCGTGGTAAAGTGGCGAAGTGTCCGCTCCGGAACACCGGAAA 600
QY 601 CCCCAATCCCGCTGCGGCGCTCTGACACCCCGCGGCGGACACGAGACTGGGCC 660
DB 601 CCCCAATCCCGCTGCGGCGCTCTGACACCCCGCGGCGGACACGAGACTGGGCC 660
QY 661 TCCGACGCGAGCGCTGCGGAGCACCGGTGCGTGGAAACCGAGGACCTTTGTAAAC 720
DB 661 TCCGACGCGAGCGCTGCGGAGCACCGGTGCGTGGAAACCGAGGACCTTTGTAAAC 720
QY 721 GCCAGTGTGCTCTTTTGAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 780
DB 721 GCCAGTGTGCTCTTTTGAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 780
QY 781 TGCGCCCTAAAGATGTCTGGGTGACTTAGATGCTAGGATCAGTTTGTCTTCAATGTAAA 840
DB 781 TGCGCCCTAAAGATGTCTGGGTGACTTAGATGCTAGGATCAGTTTGTCTTCAATGTAAA 840
QY 841 TGACACGCGGAGCTCCGTAAGGAGCTAGCAGGAGGAGCTGAAAGCGTCTTCAAGTACTGC 900
DB 841 TGACACGCGGAGCTCCGTAAGGAGCTAGCAGGAGGAGCTGAAAGCGTCTTCAAGTACTGC 900
QY 901 TGGTGGGCGGTGATGCTACAGGCGGATCAGACAGTTTGTCTTCTGGAAGTAACTGACA 960
DB 901 TGGTGGGCGGTGATGCTACAGGCGGATCAGACAGTTTGTCTTCTGGAAGTAACTGACA 960

```

```

; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 980
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-829-118-1
Query Match      100.0%; Score 980; DB 19; Length 980;
Best Local Similarity 100.0%; Pred. No. 1.7e-288;
Matches 980; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CTCCGCGGCGAAACACGCTGGGCACTCCATTCGGGGCTGTTTACTCCCAACTCTCGCGAG 60
DB 1 CTCCGCGGCGAAACACGCTGGGCACTCCATTCGGGGCTGTTTACTCCCAACTCTCGCGAG 60
QY 61 ACTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
DB 61 ACTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
QY 121 GCCCTGTCTGTCCTCCGCGCTCCCGGGCTTGGTGGCGCTCTGGAACGCGCTGGGAGCGG 180
DB 121 GCCCTGTCTGTCCTCCGCGCTCCCGGGCTTGGTGGCGCTCTGGAACGCGCTGGGAGCGG 180
QY 181 GACACGCGGCGGAGGATGGAAGAGTCTCGGACATTTGGCGGGCGGGCGGGTGG 240
DB 181 GACACGCGGCGGAGGATGGAAGAGTCTCGGACATTTGGCGGGCGGGCGGGTGG 240
QY 241 CAGGTTGAAAGCGGAGGCGTGGCCAGCGAGCTGCCAGGCGGAGAACCGCTGGGG 300
DB 241 CAGGTTGAAAGCGGAGGCGTGGCCAGCGAGCTGCCAGGCGGAGAACCGCTGGGG 300
QY 301 AACCTTTGGTCCGCTCTGCGGCTCGCTTAGATCCCGAAAAAGAGCACGGGCGGAAA 360
DB 301 AACCTTTGGTCCGCTCTGCGGCTCGCTTAGATCCCGAAAAAGAGCACGGGCGGAAA 360
QY 361 GCGGCGAGCTGGGCGAGGATCTAGAAAGACTGCTGGCGGAGGCTCCCTGCCCGCGG 420
DB 361 GCGGCGAGCTGGGCGAGGATCTAGAAAGACTGCTGGCGGAGGCTCCCTGCCCGCGG 420
QY 421 GCCTGTCTCATGACTCGTGGAGCTCGCTTCCCGCGGAGCCCTTCTCTGCAAGGGTC 480
DB 421 GCCTGTCTCATGACTCGTGGAGCTCGCTTCCCGCGGAGCCCTTCTCTGCAAGGGTC 480
QY 481 CACGTCCAGGACCGGCGGCTCGGACACCCACCCCGCGGCGGACCTGCTGCGGTGGTGC 540
DB 481 CACGTCCAGGACCGGCGGCTCGGACACCCACCCCGCGGCGGACCTGCTGCGGTGGTGC 540
QY 541 CCTTTAAACCCCGGCGTGGTAAAGTGGCGAAGTGTCCGCTCCGGAACACCGGAAA 600
DB 541 CCTTTAAACCCCGGCGTGGTAAAGTGGCGAAGTGTCCGCTCCGGAACACCGGAAA 600
QY 601 CCCCAATCCCGCTGCGGCGCTCTGACACCCCGCGGCGGACACGAGACTGGGCC 660
DB 601 CCCCAATCCCGCTGCGGCGCTCTGACACCCCGCGGCGGACACGAGACTGGGCC 660
QY 661 TCCGACGCGAGCGCTGCGGAGCACCGGTGCGTGGAAACCGAGGACCTTTGTAAAC 720
DB 661 TCCGACGCGAGCGCTGCGGAGCACCGGTGCGTGGAAACCGAGGACCTTTGTAAAC 720
QY 721 GCCAGTGTGCTCTTTTGAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 780
DB 721 GCCAGTGTGCTCTTTTGAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 780
QY 781 TGCGCCCTAAAGATGTCTGGGTGACTTAGATGCTAGGATCAGTTTGTCTTCAATGTAAA 840
DB 781 TGCGCCCTAAAGATGTCTGGGTGACTTAGATGCTAGGATCAGTTTGTCTTCAATGTAAA 840
QY 841 TGACACGCGGAGCTCCGTAAGGAGCTAGCAGGAGGAGCTGAAAGCGTCTTCAAGTACTGC 900
DB 841 TGACACGCGGAGCTCCGTAAGGAGCTAGCAGGAGGAGCTGAAAGCGTCTTCAAGTACTGC 900
QY 901 TGGTGGGCGGTGATGCTACAGGCGGATCAGACAGTTTGTCTTCTGGAAGTAACTGACA 960
DB 901 TGGTGGGCGGTGATGCTACAGGCGGATCAGACAGTTTGTCTTCTGGAAGTAACTGACA 960

```

```

RESULT 2
US-10-829-118-1
; Sequence 1, Application US/10829118
; Publication No. US20040191759A1
; GENERAL INFORMATION:
; APPLICANT: Pharmacia AB
; TITLE OF INVENTION: Promoter Sequences
; FILE REFERENCE: 00130
; CURRENT APPLICATION NUMBER: US/10/829,118
; PRIORITY FILING DATE: 2004-04-20
; PRIOR APPLICATION NUMBER: US/09/891,711
; PRIOR FILING DATE: 2001-06-26

```

Db 901 TGGTGGCGGTGATCGCTACAGCCGATCAGACAGTTTGTGCTCTGGAACCTTGACA 960
QY 961 CTGCACACCGTAAATGCTGA 980
Db 961 CTGCACACCGTAAATGCTGA 980

RESULT 3

US-09-891-711-2
; Sequence 2, Application US/09891711
; Patent No. US20020082404A1
; GENERAL INFORMATION:
; APPLICANT: Pharmacia AB
; FILE REFERENCE: 00130
; TITLE OF INVENTION: Promoter Sequences
; CURRENT APPLICATION NUMBER: US/09/891,711
; CURRENT FILING DATE: 2001-06-26
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2469
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-891-711-2

Query Match 100.0%; Score 980; DB 9; Length 2469;
Best Local Similarity 100.0%; Pred. No. 2.3e-288;
Matches 980; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCGCGGGGAACACGCTGGGCACTCCATTCGGGGCTGTTTACTCCCAACTCTCGCGAG 60
Db 353 CTCGCGGGGAACACGCTGGGCACTCCATTCGGGGCTGTTTACTCCCAACTCTCGCGAG 412
QY 61 ACTGGGCGGCGGCGGCGAGGAGCCACAGCTGGAGCCTCAGCTCGCGGAGCCAGCGT 120
Db 413 ACTGGGCGGCGGCGGCGAGGAGCCACAGCTGGAGCCTCAGCTCGCGGAGCCAGCGT 472
QY 121 GCCTCTGTCTGCTCCCGGCTTCCGGGCTTCGCGCTGCGAGCGCTGGGAGCGG 180
Db 473 GCCTCTGTCTGCTCCCGGCTTCCGGGCTTCGCGCTGCGAGCGCTGGGAGCGG 532
QY 181 GACCACCGCGGAGGATGGAAGAAGTGTCTCGACATTTTCGCGCGGCGGCGCGGCTGG 240
Db 533 GACCACCGCGGAGGATGGAAGAAGTGTCTCGACATTTTCGCGCGGCGGCGGCGGCTGG 592
QY 241 CAGGGTGGAAAGCGGAGGCGGCTGGCCAGAGCTGCGAGCGGAGCGGCGGAGCGGCTGGGG 300
Db 593 CAGGGTGGAAAGCGGAGGCGGCTGGCCAGAGCTGCGAGCGGAGCGGCGGAGCGGCTGGGG 652
QY 301 AACCCCTTGTTCGCTCTGCGCGTCTAGGATCCCGAAAGGAGGACGCGGCGGAAA 360
Db 653 AACCCCTTGTTCGCTCTGCGCGTCTAGGATCCCGAAAGGAGGACGCGGCGGAAA 712
QY 361 GCGGCGAGGCTGGGCGAGGATCTAGAAAGACTGCTTGGGCGAGGCTCCCTGCCCCCGCGG 420
Db 713 GCGGCGAGGCTGGGCGAGGATCTAGAAAGACTGCTTGGGCGAGGCTCCCTGCCCCCGCGG 772
QY 421 GCCTGTGTATGGAATCTGTGTGAGAGCTGCTTCCCGCGCGGAGCCCTTCTCAGGGGTC 480
Db 773 GCCTGTGTATGGAATCTGTGTGAGAGCTGCTTCCCGCGCGGAGCCCTTCTCAGGGGTC 832
QY 481 CAGTCCAGGACCGGGGCTCGGACACCCACCCCGGCGGCGAGCTGCCCTGGGTGC 540
Db 833 CAGTCCAGGACCGGGGCTCGGACACCCACCCCGGCGGCGAGCTGCCCTGGGTGC 592
QY 541 CCCTTAACCCGGGCGGTAGTCTGTTAAGATGGCGAAGTGTCCGGTCCGGAAACACGCGAAA 600
Db 893 CCCTTAACCCGGGCGGTAGTCTGTTAAGATGGCGAAGTGTCCGGTCCGGAAACACGCGAAA 952
QY 601 CCCCAAATCCCGCTCCCGAGCTCTCTGACCCCGGCGGCGGAGCGAGACTGGGCC 660
Db 953 CCCCAAATCCCGCTCCCGAGCTCTCTGACCCCGGCGGCGGAGCGAGACTGGGCC 1012

QY 661 TCCCGACGCGGACGCGGCTGCCGGGACACCGGTGCGTGGGAAACGGAGGACCTTTGTAAAC 720
Db 1013 TCCCGACGCGGACGCGGCTGCCGGGACACCGGTGCGTGGGAAACGGAGGACCTTTGTAAAC 1072
QY 721 GCCACGTGTTGCTCTTTTGTGAAAGAAACAAAGATAAATGTGTTAAACTGTCTGAAAGCT 780
Db 1073 GCCACGTGTTGCTCTTTTGTGAAAGAAACAAAGATAAATGTGTTAAACTGTCTGAAAGCT 1132
QY 781 TGCCGCTTAAAGATGCTCTGGGTGACTTAGATGCTAGGATCAGTTTGTGTTTCAATGTAAA 840
Db 1133 TGCCGCTTAAAGATGCTCTGGGTGACTTAGATGCTAGGATCAGTTTGTGTTTCAATGTAAA 1192
QY 841 TGGACCAAGCCCGGACTCCGTACGGCACTAGCAGGGGACTGAAAGCGCTTTCAGGTACTGC 900
Db 1193 TGGACCAAGCCCGGACTCCGTACGGCACTAGCAGGGGACTGAAAGCGCTTTCAGGTACTGC 1252
QY 901 TGGTGGGCGGTGATGCGCTACAGGCGCGATCAGACAGTTTGTGCTCTTGGAACTTGACA 960
Db 1253 TGGTGGGCGGTGATGCGCTACAGGCGCGATCAGACAGTTTGTGCTCTTGGAACTTGACA 1312
QY 961 CTGCACACCGTAAATGCTGA 980
Db 1313 CTGCACACCGTAAATGCTGA 1332

RESULT 4

US-10-829-118-2
; Sequence 2, Application US/10829118
; Publication No. US20040191759A1
; GENERAL INFORMATION:
; APPLICANT: Pharmacia AB
; TITLE OF INVENTION: Promoter Sequences
; FILE REFERENCE: 00130
; CURRENT APPLICATION NUMBER: US/10/829,118
; PRIOR FILING DATE: 2004-04-20
; PRIOR APPLICATION NUMBER: US/09/891,711
; PRIOR FILING DATE: 2001-06-26
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2469
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-829-118-2

Query Match 100.0%; Score 980; DB 19; Length 2469;
Best Local Similarity 100.0%; Pred. No. 2.3e-288;
Matches 980; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCGCGGGGAACACGCTGGGCACTCCATTCGGGGCTGTTTACTCCCAACTCTCGCGAG 60
Db 353 CTCGCGGGGAACACGCTGGGCACTCCATTCGGGGCTGTTTACTCCCAACTCTCGCGAG 412
QY 61 ACTGGGCGGCGGCGGCGAGGCGCCACAGCTGGAGCCTCAGCTCCGCGGAGCCAGCGT 120
Db 413 ACTGGGCGGCGGCGGCGAGGCGCCACAGCTGGAGCCTCAGCTCCGCGGAGCCAGCGT 472
QY 121 GCCTCTGTCTGCTCCCGGCTTCCGGGCTTCGCGCTGCGCGCTCTGGAAGCGGCTGGGAGCGG 180
Db 473 GCCTCTGTCTGCTCCCGGCTTCCGGGCTTCGCGCTGCGCGCTCTGGAAGCGGCTGGGAGCGG 532
QY 181 GACCACCGCGGAGGATGGAAGAAGTGTCTCGGACATTTTCGCGCGGCGGCGGCGGCTGG 240
Db 533 GACCACCGCGGAGGATGGAAGAAGTGTCTCGGACATTTTCGCGCGGCGGCGGCGGCTGG 592
QY 241 CAGGGTGGAAAGCGGAGGCGGCTGGCCAGAGCTGCGAGCGGCGGAGGAGCGGCTGGGG 300
Db 593 CAGGGTGGAAAGCGGAGGCGGCTGGCCAGAGCTGCGAGCGGCGGAGGAGCGGCTGGGG 652
QY 301 AACCCCTTGTTCGCTCTGCGCGTCTGCGCGTCTAGGATCCCGAAAGGAGGACGCGGCGGAAA 360
Db 653 AACCCCTTGTTCGCTCTGCGCGTCTGCGCGTCTAGGATCCCGAAAGGAGGACGCGGCGGAAA 712

```
QY 361 GCGGCCAGGCTGGGACAGATCTAGAAAGACTGCCTTGGCGAGGCTCCCTGCCCCCGCG 420
|
|
|
Db 713 GCGGCCAGGCTGGGACAGATCTAGAAAGACTGCCTTGGCGAGGCTCCCTGCCCCCGCG 772
|
|
|
QY 421 GCCTGCTGCATGGAAGCTGGAGAGCTCGCTTCCCGCGGAGACCTTCTCTCAGGGTTC 480
|
|
|
Db 773 GCCTGCTGCATGGAAGCTGGAGAGCTCGCTTCCCGCGGAGACCTTCTCTCAGGGTTC 832
|
|
|
QY 481 CACGTCCAGGCACCGCGGCTCGGACACCCCAACCCCGCGGACCTGCCCTCGGGTGC 540
|
|
|
Db 833 CACGTCCAGGCACCGCGGCTCGGACACCCCAACCCCGCGGACCTGCCCTCGGGTGC 892
|
|
|
QY 541 CCTTTAACCCGGCGGTAGCTCGTTAAGATGGCGAAGTGTCCGTCGGGAAACACGCGAAA 600
|
|
|
Db 893 CCTTTAACCCGGCGGTAGCTCGTTAAGATGGCGAAGTGTCCGTCGGGAAACACGCGAAA 952
|
|
|
QY 601 CCCCCAATCCCGCTGCCGACCTCTGTACCCCGCGGCCCAACCGGACGACAGCTGGGCC 660
|
|
|
Db 953 CCCCCAATCCCGCTGCCGACCTCTGTACCCCGCGGCCCAACCGGACGACAGCTGGGCC 1012
|
|
|
QY 661 TCCCGACGCGCAGCGCGCTGCCGGGACACCGGTGCGTGGGAAACCGGAGACCTTTGTAAC 720
|
|
|
Db 1013 TCCCGACGCGCAGCGCGCTGCCGGGACACCGGTGCGTGGGAAACCGGAGACCTTTGTAAC 1072
|
|
|
QY 721 GCCAGGTGTTGCTCTTTTGGAAAAAACAAGATAAATGTGTTAAACTGTCTGAAAAGCT 780
|
|
|
Db 1073 GCCAGGTGTTGCTCTTTTGGAAAAAACAAGATAAATGTGTTAAACTGTCTGAAAAGCT 1132
|
|
|
QY 781 TCGCGCCTAAAGATGCTGGGTGACTTAGATGCTAGGATCAGTTGTTTCAATGTAAA 840
|
|
|
Db 1133 TCGCGCCTAAAGATGCTGGGTGACTTAGATGCTAGGATCAGTTGTTTCAATGTAAA 1192
|
|
|
QY 841 TGGACACGCGGACCTCCGTACGGCACTAGCAGGGGACTGAAAGCGCTTTCAGGTACTGC 900
|
|
|
Db 1193 TGGACACGCGGACCTCCGTACGGCACTAGCAGGGGACTGAAAGCGCTTTCAGGTACTGC 1252
|
|
|
QY 901 TGGTGGCGGTGATGCGCTACAGCGCGATCAGACAGTTTGTCTTCTGGAACCTTGACA 960
|
|
|
Db 1253 TGGTGGCGGTGATGCGCTACAGCGCGATCAGACAGTTTGTCTTCTGGAACCTTGACA 1312
|
|
|
QY 961 CTGCACACGGTAATCTCTGA 980
|
|
|
Db 1313 CTGCACACGGTAATCTCTGA 1332
|
|
|
RESULT 5
US-09-891-711-3/c
; Sequence 3, Application US/09891711
; Patent No. US20020082404A1
; GENERAL INFORMATION:
; APPLICANT: Pharmacia AB
; TITLE OF INVENTION: Promoter Sequences
; FILE REFERENCE: 00130
; CURRENT APPLICATION NUMBER: US/09/891,711
; CURRENT FILING DATE: 2001-06-26
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 4338
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (497)..(3655)
US-09-891-711-3
Query Match 12.7%; Score 124.4; DB 9; Length 4338;
Best Local Similarity 99.2%; Pred. No. 5.4e-27;
Matches 125; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CTCCCGGCGGAACACCGCTGGGCATCCATTTCGGGGCTGTTTACTCCCAACTCTCGCGAG 60
|
|
|
Db 126 CTCCCGGCGGAACACCGCTGGGCATCCATTTCGGGGCTGTTTACTCCCAACTCTCGCGAG 67
|
|
|
RESULT 6
US-10-172-118-953/c
; Sequence 953, Application US/10172118
; Publication No. US20030224374A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Chris
; APPLICANT: Van 't Veer, Laura
; APPLICANT: Van de Vijver, Marc
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-175-999
; CURRENT APPLICATION NUMBER: US/10/172,118
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/380,770
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 953
; LENGTH: 4338
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NM_003791
; DATABASE ENTRY DATE: 2001-06-18
US-10-172-118-953
Query Match 12.7%; Score 124.4; DB 17; Length 4338;
Best Local Similarity 99.2%; Pred. No. 5.4e-27;
Matches 125; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CTCCCGGCGGAACACCGCTGGGCATCCATTTCGGGGCTGTTTACTCCCAACTCTCGCGAG 60
|
|
|
Db 126 CTCCCGGCGGAACACCGCTGGGCATCCATTTCGGGGCTGTTTACTCCCAACTCTCGCGAG 67
|
|
|
QY 61 ACTGGGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
|
|
|
Db 66 ACTGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
|
|
|
QY 121 GCCCTG 126
|
|
|
Db 6 GCCCTG 1
|
|
|
RESULT 7
US-10-342-887-953/c
; Sequence 953, Application US/10342887
; Publication No. US20040058340A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter S.
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Christopher J.
; APPLICANT: Van 't Veer, Laura Johanna
; APPLICANT: Van de Vijver, Marc J.
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-198-999
; CURRENT APPLICATION NUMBER: US/10/342,887
```



```

; CURRENT FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/298,918
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/380,710
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 10/172,118
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 953
; LENGTH: 4338
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-342-887-953

```

```
Query Match      12.7%; Score 124.4; DB 18; Length 4338;
Best Local Similarity 99.2%; Pred. NO. 5.4e-27;
Matches 125; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

1 CTCGGGGGGAACAGCGCTGGGCACTCATTCGGGGCTGTTTACTCCCAACTCTCCGAG 60
126 CTCGGGGGGAACAGCGCTGGGCACTCATTCGGGGCTGTTTACTCCCAACTCTCCGAG 67
61 ACTGGCGGCGGGCGAGCGAGGGCCACAGCTGGGGCCCTCAGCTCGCGCGACCCAGCT 120
65 ACTGGCGCACGGGGCCAGCGAGGGCCACAGCTGGGGACCTCAGCTCGCGACCCAGCT 7

```
Query Match      12.7%; Score 124.4; DB 19; Length 4338;
Best Local Similarity 99.2%; Pred. No. 5.4e-27;
Matches 125; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

61	ACTGGCGCGCGCGGCGACGAGGCCCA	CAGCTGGAGCCTCAGCTCGCGGACCCAGCGT	120
Qy			
Db			
66	ACTGGCGCACCGGGCCAGCGAGGCCCA	CAGCTGGAGCCTCAGCTCGCGGACCCAGCGT	7
Qy			
Db			
121	GCCCTG	126	
Qy			
Db			
6	GCCCTG	1	
Qy			
Db			

; Sequence 1116, Application US/10956157
 ; Publication No. US20050118625A1
 ; GENERAL INFORMATION:

	Query Match	12.7%;	Score 124.4;	DB 21;	Length 4338;
	Best Local Similarity	99.2%;	Pred. No. 5.4e-27;		
	Matches 125;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
QY	1	CTCCGGCGGACACGCCCTGGGCACCTCCATTCTGGGGCTGTTTACTCCCAACTCTCCGCGAG	60		
Db	126	CTCCGGCGGGAACACGCCCTGGGCACCTCCATTCTGGGGCTGTTTACTCCCAACTCTCCGCGAG	67		
QY	61	ACTGGCGCGCCGGCGGCAGCGAGGCCACACAGCTGGAGCCTCAGCTCCGGCGAGCCCAAGCGT	120		
Db	66	ACTGGCGGACCGGGCCAGCGAGGCCACACAGCTGGAGCCTCAGCTCCGGCGAGCCCAAGCGT	7		
QY	121	GCCCTG 126			
Db	6	GCCCTG 1			

Query Match	5.8%	Score 56.4;	DB 19;	Length 1523;
Best Local Similarity	51.6%;	Pred. No. 2.2e-06;		
Matches 123;	Conservative	0;	Mismatches 121;	Indels 0;
			Gaps	0;

193	AGGATGGACAAAGGTGCTCGCGACATTTTCGGCGCGCGGGGCGCGTGGCAGGCGTGAAGC	252
QY		
219	GGACCGCGGGAGGAGCCCGCGGCTCGAGGCGCGGCGCGGCGCCGAGCGCGCGG	160
Db		
253	CGAGGGGGGTGGCCAGCAGCTGCCAGGCGGCGAGAACCGCTGGGGGACCCCTTGGTCC	312
QY		
159	GGGCGCGCGCGCGCGCGGGGGGAGGCGCGGAAACGGGCGGCGAGGACCGGCGACG	100
Db		

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	980	100.0	980	4	US-09-891-711-1	Sequence 1, Appli
2	980	100.0	2469	4	US-09-891-711-2	Sequence 2, Appli
C 3	966.4	98.6	67156	4	US-09-949-016-12884	Sequence 12284, A
C 4	966.4	98.6	67157	4	US-09-949-016-16558	Sequence 16558, A
C 5	126	12.9	4338	4	US-09-949-016-4816	Sequence 4816, Ap
C 6	124.4	12.7	4338	3	US-09-360-237-4	Sequence 4, Appli
C 7	124.4	12.7	4338	4	US-09-891-711-3	Sequence 3, Appli
C 8	124.4	12.7	4338	4	US-09-949-016-542	Sequence 542, App
C 9	53	5.4	4403765	3	US-09-103-840A-2	Sequence 2, Appli
C 10	52.2	5.3	300402	4	US-09-949-016-13632	Sequence 13632, A
C 11	48	4.9	42053	4	US-09-949-016-15924	Sequence 15924, A
C 12	47.4	4.8	154746	4	US-09-827-688-8	Sequence 8, Appli
C 13	47.4	4.8	154746	4	US-09-827-688-8	Sequence 8, Appli
C 14	46.8	4.8	2571	4	US-09-984-880-1	Sequence 1, Appli
C 15	46.8	4.8	2571	4	US-10-377-032-1	Sequence 1, Appli
C 16	46.8	4.8	20966	4	US-09-984-880-3	Sequence 3, Appli
C 17	46.8	4.8	20966	4	US-10-377-032-3	Sequence 3, Appli
C 18	46.6	4.8	1736	3	US-09-162-524-2	Sequence 2, Appli
C 19	46	4.7	867	4	US-09-302-540-9534	Sequence 9534, Ap
C 20	46	4.7	9867	4	US-09-302-540-1008	Sequence 1008, Ap
C 21	45.8	4.7	7218	1	US-08-232-463-14	Sequence 14, Appl
C 22	45.8	4.7	128516	4	US-09-949-016-13501	Sequence 13501, A
C 23	45.8	4.7	133157	4	US-09-949-016-12541	Sequence 12541, A
C 24	45.4	4.6	4335	3	US-08-974-549A-6	Sequence 6, Appli
C 25	45.4	4.6	4335	4	US-09-721-456-6	Sequence 6, Appli
C 26	45.2	4.6	4048	4	US-09-976-594-1053	Sequence 1053, Ap
C 27	45.2	4.6	122772	4	US-09-949-016-14132	Sequence 14132, A

```

QY 421 GCCTGCTGTATGAGTCTGTGAGAGCTCGCTTCCGCGCGGACCCCTTCCTGACGGGTC 480
Db 421 GCCTGCTGTATGAGTCTGTGAGAGCTCGCTTCCGCGCGGACCCCTTCCTGACGGGTC 480
QY 481 CAGGTCCAGGACCCGCGCTCGGACACCCACCCCGCGCGGACCTGCGCTCGGTCG 540
Db 481 CAGGTCCAGGACCCGCGCTCGGACACCCACCCCGCGCGGACCTGCGCTCGGTCG 540
QY 541 CCCTTAAACCCGCGCGGTAGCTGTTAAGATGCGGAAGTGTCCGCTCCGGAACACCGGAAA 600
Db 541 CCCTTAAACCCGCGCGGTAGCTGTTAAGATGCGGAAGTGTCCGCTCCGGAACACCGGAAA 600
QY 601 CCCCCAAATCCCGCTTCCCGACCTCTGTACCCCGCGCGGACCGGACGACGACTGGGCC 660
Db 601 CCCCCAAATCCCGCTTCCCGACCTCTGTACCCCGCGCGGACCGGACGACGACTGGGCC 660
QY 661 TCCGACGCGGACCGGCTGCGGACACCCGCTGCGTGGGAACCGGAGACCTTTGTAAC 720
Db 661 TCCGACGCGGACCGGCTGCGGACACCCGCTGCGTGGGAACCGGAGACCTTTGTAAC 720
QY 721 GCCAGGTGTTGCTCTTTTGAAGAAACAAAGATTAATGTTTAAACTGTCTGAAAAGCT 780
Db 721 GCCAGGTGTTGCTCTTTTGAAGAAACAAAGATTAATGTTTAAACTGTCTGAAAAGCT 780
QY 781 TCCGCGCTTAAAGATGTTGGGTGACTTAGATGCTAGGATCAGTTGTTTCAATGTAAA 840
Db 781 TCCGCGCTTAAAGATGTTGGGTGACTTAGATGCTAGGATCAGTTGTTTCAATGTAAA 840
QY 841 TGGACACGCGGACTCCGTACGCGACTAGCAGGGGACTGAAAGCGTCTTCAGGTACTGC 900
Db 841 TGGACACGCGGACTCCGTACGCGACTAGCAGGGGACTGAAAGCGTCTTCAGGTACTGC 900
QY 901 TGGTGGGCGGTGATGCGGTACAGGCGGATCAGACAGATTTGTGTCTTCTGGAACCTTGACA 960
Db 901 TGGTGGGCGGTGATGCGGTACAGGCGGATCAGACAGATTTGTGTCTTCTGGAACCTTGACA 960
QY 961 CTGCACCGGTAATGCTGA 980
Db 961 CTGCACCGGTAATGCTGA 980

```

RESULT 2

```

US-09-891-711-2
; Sequence 2, Application US/09891711
; Patent No. 6723553
; GENERAL INFORMATION:
; APPLICANT: Pharmacia AB
; TITLE OF INVENTION: Promoter Sequences
; FILE REFERENCE: 00130
; CURRENT APPLICATION NUMBER: US/09/891,711
; CURRENT FILING DATE: 2001-06-26
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2469
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-891-711-2

```

```

Query Match 100.0%; Score 980; DB 4; Length 2469;
Best Local Similarity 100.0%; Pred. No. 4.4e-250;
Matches 980; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CTCGCGGGGAAACAGCCCTGGGCACTCCATTCGGGCTGTTTACTCCCAACTCTCGCGAG 60
Db 353 CTCGCGGGGAAACAGCCCTGGGCACTCCATTCGGGCTGTTTACTCCCAACTCTCGCGAG 412
QY 61 ACTGGCGCGCGGCGAGCGGCGGACAGTCTGAGTCCGCGGACCCAGCGT 120
Db 413 ACTGGCGCGCGGCGAGCGGCGGACAGTCTGAGTCCGCGGACCCAGCGT 472
QY 121 GCCTGTCTGTCCGCGCTCCGCGGCTTGGTGGCGCTCTGGAGCGGCTGGGCGG 180

```

```

Db 473 GCCCTGTCTGTCCCGCGCTCCCGGGCTTCGTGCGCGCTCTGGACGCGCTGGGCGAGCGG 532
QY 181 GACCAACCGCGGAGGATGGAAGAGTCTCCGACATTTGCGGCGGCGGCGGCGCGGTGG 240
Db 533 GACCAACCGCGGAGGATGGAAGAGTCTCCGACATTTGCGGCGGCGGCGGCGCGGTGG 592
QY 241 CAGGTTGAAGCGGAGGGCGTGGCCAGGAGCTGCCAGGCGGCGGAGAACCGGCTGGGGG 300
Db 593 CAGGTTGAAGCGGAGGGCGTGGCCAGGAGCTGCCAGGCGGCGGAGAACCGGCTGGGGG 652
QY 301 AACCTTTCGCTCCGCTCTCGCGTCTAGGATCCCGGAAAGAGACACGCGCGCGAAA 360
Db 653 AACCTTTCGCTCTCGCGTCTAGGATCCCGGAAAGAGACACGCGCGCGAAA 712
QY 361 CGGCGCAGGCTGGGCGCAGGATCTAGAAAGACTGCTTGGCGCAGGCTCCCTGCCCCCGGG 420
Db 713 GCGGCGCAGGCTGGGCGCAGGATCTAGAAAGACTGCTTGGCGCAGGCTCCCTGCCCCCGGG 772
QY 421 GCCTGCTGTATGAGTCTGTGAGAGCTCGCTTCCGCGCGGACCCCTTCCTGACGGGTC 480
Db 773 GCCTGCTGTATGAGTCTGTGAGAGCTCGCTTCCGCGCGGACCCCTTCCTGACGGGTC 832
QY 481 CAGGTCCAGGACCGCGGCTCGGACACCCACCCCGCGCGGACCTTCGCGCTGGGTGC 540
Db 833 CAGGTCCAGGACCGCGGCTCGGACACCCACCCCGCGCGGACCTTCGCGCTGGGTGC 892
QY 541 CCCTTAAACCCGCGCGGTAGCTGTTAAGATGCGGAAGTGTCCGTCGGAACACCGGAAA 600
Db 893 CCCTTAAACCCGCGCGGTAGCTGTTAAGATGCGGAAGTGTCCGTCGGAACACCGGAAA 952
QY 601 CCGCAATCCCGCTTCCCGGACCTCTGACCCCGCGCGGACCGGACGACGACTGGGCG 660
Db 953 CCGCAATCCCGCTTCCCGGACCTCTGACCCCGCGCGGACCGGACGACGACTGGGCG 1012
QY 661 TCCGACGCGGACCGGCTGCGGACACCCGCTGCGTGGCGGAAACCGAGAGACCTTTGTAAC 720
Db 1013 TCCGACGCGGACCGGCTGCGGACACCCGCTGCGTGGCGGAAACCGAGGACCTTTGTAAC 1072
QY 721 GCCAGCTGTTGCTCTTTTGAAGAAACAAAGATTAATGTTTAAACTGTCTGAAAAGCT 780
Db 1073 GCCAGCTGTTGCTCTTTTGAAGAAACAAAGATTAATGTTTAAACTGTCTGAAAAGCT 1132
QY 781 TGCCGCTTAAAGATGTTGGGTGACTTAGATGCTAGGATCAGTTGTTTCAATGTAAA 840
Db 1133 TGCCGCTTAAAGATGTTGGGTGACTTAGATGCTAGGATCAGTTGTTTCAATGTAAA 1192
QY 841 TGGACACCGCGGACTCCGTACGCGCACTAGCAGGGGACTGAAAGCGTCTTCAGGTACTGC 900
Db 1193 TGGACACCGCGGACTCCGTACGCGCACTAGCAGGGGACTGAAAGCGTCTTCAGGTACTGC 1252
QY 901 TGGTGGGCGGTGATGCGCTTACAGGCGGATCAGACAGATTTGTGTCTTCTGGAACCTTGACA 960
Db 1253 TGGTGGGCGGTGATGCGCTTACAGGCGGATCAGACAGATTTGTGTCTTCTGGAACCTTGACA 1312
QY 961 CTGCACCGGTAATGCTGA 980
Db 1313 CTGCACCGGTAATGCTGA 1332

```

RESULT 3

```

US-09-949-016-12284/c
; Sequence 12284, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768

```

;
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12284
; LENGTH: 67156
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12284

Query Match 98.6%; Score 966.4; DB 4; Length 67156;
Best Local Similarity 99.8%; Pred. No. 6.2e-246;
Matches 978; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 CTCGCGGGAACACGCTGGGCACTCCATTGGGGCTGTTTACTCCCAACTCTCGCGAG 60
DB 2126 CTCGCGGGAACACGCTGGGCACTCCATTGGGGCTGTTTACTCCCAACTCTCGCGAG 2067

QY 61 ACTGGCGCGCGGCGGAGGCGCCACAGCTGGGAGCCTCAGCTCCGCGGACCCAGCGT 120
DB 2066 ACTGGCGCGCGGCGGAGGCGCCACAGCTGGGAGCCTCAGCTCCGCGGACCCAGCGT 2007

QY 121 GCCCTGTCTGTCCCGCGCTCCCGGGCTTTCGCTGCGCGCTCTGGAGCGCGTGAGCGG 180
DB 2006 GCCCTGTCTGTCCCGCGCTCCCGGGCTTTCGCTGCGCGCTCTGGAGCGCGTGAGCGG 1947

QY 181 GACACGCGCGGAGGATGACCAAGGTGCTCCGACATTTGCGGCGCGGGCGCGGTGG 240
DB 1946 GACACGCGCGGAGGATGACCAAGGT-CTCGACATTTGCGGCGCGGGCGCGGTGG 1888

QY 241 CAGGGTGAAGCGGAGGCGCTGGCCAGCAGCTGCGAGCGCGGAGGAGCGGCGTGGGG 300
DB 1887 CAGGGTGAAGCGGAGGCGCTGGCCAGCAGCTGCGAGCGCGGAGGAGCGGCGTGGGG 1828

QY 301 AACCTTGTCTGCTCTGCGCGCTCTGAGGATCCCGAAAGGAGCAGCGGCGCGAAA 360
DB 1827 AACCTTGTCTGCTCTGCGCGCTCTGAGGATCCCGAAAGGAGCAGCGGCGCGAAA 1768

QY 361 GCGGCGAGCTGGGCGAGATCTAGAAAGCTGCTGGGCGGAGGCTCCCTGCGCGCGCG 420
DB 1767 GCGGCGAGCTGGGCGAGATCTAGAAAGCTGCTGGGCGGAGGCTCCCTGCGCGCGCG 1708

QY 421 GCCTGTCTGATGAGCTCTGAGAGCTCGCTTCCCGCGGAGCCTTCTCTGAGGGGTC 480
DB 1707 GCCTGTCTGATGAGCTCTGAGAGCTCGCTTCCCGCGGAGCCTTCTCTGAGGGGTC 1648

QY 481 CAGCTCCAGGACCGGCGGCTCGGACACCCCGGCGGCGGAGCCTGCGCTGGGGTGC 540
DB 1647 CAGCTCCAGGACCGGCGGCTCGGACACCCCGGCGGCGGAGCCTGCGCTGGGGTGC 1588

QY 541 CCCTTAAACCGGCGGCTAGCTGTTAAGATGGGGAAGTGTCCGCTCCGGAACACGCGAAA 600
DB 1587 CCCTTAAACCGGCGGCTAGCTGTTAAGATGGGGAAGTGTCCGCTCCGGAACACGCGAAA 1528

QY 601 CCCCAATCCCGCTGCCGAGCTCTGACCCCGCGGCGGAGGAGCAGACTGGGCG 660
DB 1527 CCCCAATCCCGCTGCCGAGCTCTGACCCCGCGGCGGAGGAGCAGACTGGGCG 1468

QY 661 TCCCGACGCGGCGGCTGCCGGAACACCGGTGCGTGGGAAAGGAGGAGCCTTTGTAAAC 720
DB 1467 TCCCGACGCGGCGGCTGCCGGAACACCGGTGCGTGGGAAAGGAGGAGCCTTTGTAAAC 1408

QY 721 GCCAGCTGTTGCTCTTTTGTAAAGAAACAAAGATAAATGTGTAACTCTCTGAAAGCT 780
DB 1407 GCCAGCTGTTGCTCTTTTGTAAAGAAACAAAGATAAATGTGTAACTCTCTGAAAGCT 1348

QY 781 TCGCGCTTAAAGATGCTGGGTGACTTAGATGCTAGGATCAGTTGTTTCAATGTAAA 840
DB 1347 TCGCGCTTAAAGATGCTGGGTGACTTAGATGCTAGGATCAGTTGTTTCAATGTAAA 1288

QY 841 TGGACCGCGGAGCTCCCGTACGGCACTAGCGGGAGCTGAAAGCGTCTTCAGGTACTGC 900

DB 1287 TGGACACGCGCGGACTCCCGTACGGCACTAGACGGGAGCTGAAAGCGCTTTCAGGTACTGC 1228
QY 901 TGGTGGGCGGTGATCGCTTACAGGCGGATCAGACAGTCTTGTCTCTTCGAACTTGACA 960
DB 1227 TGGTGGGCGGTGATCGCTTACAGGCGGATCAGACAGTCTTGTCTCTTCGAACTTGACA 1168
QY 961 CTGCACCAACGCTTAATGCTGA 980
DB 1167 CTGCACCAACGCTTAATGCTGA 1148

RESULT 4
US-09-949-016-16558/c
; Sequence 16558, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIORITY FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16558
; LENGTH: 67157
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16558

Query Match 98.6%; Score 966.4; DB 4; Length 67157;
Best Local Similarity 99.8%; Pred. No. 6.2e-246;
Matches 978; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 CTCGCGGGAACACGCTGGGCACTCCATTGGGGCTGTTTACTCCCAACTCTCGCGAG 60
DB 2126 CTCGCGGGAACACGCTGGGCACTCCATTGGGGCTGTTTACTCCCAACTCTCGCGAG 2067

QY 61 ACTGGCGCGCGGCGGAGGCGCCACAGCTGGGAGCCTCAGCTCCGCGGACCCAGCGT 120
DB 2066 ACTGGCGCGCGGCGGAGGCGCCACAGCTGGGAGCCTCAGCTCCGCGGACCCAGCGT 2007

QY 121 GCCTGTCTGTCTCCCGCGCTTCCCGGGCTTTCGCTGCGCGCTCTGGAGCGCGTGAGCGG 180
DB 2006 GCCTGTCTGTCTCCCGCGCTTCCCGGGCTTTCGCTGCGCGCTCTGGAGCGCGTGAGCGG 1947

QY 181 GACACGCGCGGAGGATGACCAAGGTGCTCTCGCAATTTGCGGCGCGGGCGCGGTGG 240
DB 1946 GACACGCGCGGAGGATGACCAAGGT-CTCGCGACATTTGCGGCGCGGGCGCGGTGG 1888

QY 241 CAGGGTGAAGCGGAGGCGCTGGCCAGCAGCTGCCAGCGCGGAGGAGCAGCTGGGG 300
DB 1887 CAGGGTGAAGCGGAGGCGCTGGCCAGCAGCTGCCAGCGCGGAGGAGCAGCTGGGG 1828

QY 301 AACCTTGTCTGCTCTGCGCGCTCTAGGATCCCGAAAGGAGCAGCGGCGCGAAA 360
DB 1827 AACCTTGTCTGCTCTGCGCGCTCTAGGATCCCGAAAGGAGCAGCGGCGCGAAA 1768

QY 361 GCGGCGAGCTGGGCGAGGATCTAGAAAGCTGCTGGGCGGAGGCTCCCTGCGCGCGCG 420
DB 1767 GCGGCGAGCTGGGCGAGGATCTAGAAAGCTGCTGGGCGGAGGCTCCCTGCGCGCGCG 1708

QY 421 GCCTGTCTGATGAGCTCTGAGAGCTCGCTTCCCGCGGAGCCTTCTCTGAGGGGTC 480
DB 1707 GCCTGTCTGATGAGCTCTGAGAGCTCGCTTCCCGCGGAGCCTTCTCTGAGGGGTC 1648

QY 481 CAGCTCCAGGACCGGCGGCTCGGACACCCCGGCGGCGGAGCCTGCGCTGGGGTGC 540

Db 1647 CACGTCACGCGCGCTCGGACACCCACACCCCGCGGCACTGCCCTGGGTGC 1588
QY 541 CCCTTAACCCGCGCGGTAGCTGTTAAGATGCGGAAGTGTCCGGTCCGGAACACCGGAAA 600
Db 1587 CCCTTAACCCGCGCGGTAGCTGTTAAGATGCGGAAGTGTCCGGTCCGGAACACCGGAAA 1528
QY 601 CCCCAAAATCCCGCTGCCGACCTCTGACCCCGCGGCCCCACGAGACAGACTGGGCC 660
Db 1527 CCCCAAAATCCCGCTGCCGACCTCTGACCCCGCGGCCCCACGAGACAGACTGGGCC 1468
QY 661 TCCCGACGCGACGCGCTGCCGCGACACCGGTGCGTGCAGAAACGAGAGACCTTTGTAAAC 720
Db 1467 TCCCGACGCGACGCGCTGCCGCGACACCGGTGCGTGCAGAAACGAGAGACCTTTGTAAAC 1408
QY 721 GCCAGTGTGTCTCTTTTGAAGAAACAAAGATTAATGTGTAAACTGTCTGAAAGCT 780
Db 1407 GCCAGTGTGTCTCTTTTGAAGAAACAAAGATTAATGTGTAAACTGTCTGAAAGCT 1348
QY 781 TGCCGCTAAAGATGTCGGGTGACTTAGATGCTAGGATCAGTTTGTTCATGTAA 840
Db 1347 TGCCGCTAAAGATGTCGGGTGACTTAGATGCTAGGATCAGTTTGTTCATGTAA 1288
QY 841 TGGACACGCGGACTCCGTCAGGCACTAGCAGGGGACTGAAAGCGTCTTCAGGTACTGC 900
Db 1287 TGGACACGCGGACTCCGTCAGGCACTAGCAGGGGACTGAAAGCGTCTTCAGGTACTGC 1228
QY 901 TGGTGGGCGGTGATCGGTACAGGCGGATCAGACAGTTTGTCTCTGGAACCTTGACA 960
Db 1227 TGGTGGGCGGTGATCGGTACAGGCGGATCAGACAGTTTGTCTCTTGGAACTTGACA 1168
QY 961 CTGACACGCGTAAATGCTGA 980
Db 1167 CTGACACGCGTAAATGCTGA 1148

RESULT 5
US-09-949-016-4816/c
; Sequence 4816, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4816
; LENGTH: 4338
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-4816

Query Match 12.9%; Score 126; DB 4; Length 4338;
Best Local Similarity 100.0%; Pred. No. 2.4e-23;
Matches 126; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCGCGGCGGAACAGCGCTGGGCACTCCATTCCGGGCGCTGTTTACTCCCAACTCTCGCGAG 60
Db 126 CTCGCGGCGGAACAGCGCTGGGCACTCCATTCCGGGCGCTGTTTACTCCCAACTCTCGCGAG 67
QY 61 ACTGGGCGCGGCGGCGAGGCCACACAGCTGGAGGCTCAGCTCCGCGGACCCAGCGT 120
Db 66 ACTGGGCGCGGCGGCGAGGCCACACAGCTGGAGGCTCAGCTCCGCGGACCCAGCGT 7

QY 121 GCCCTG 126
Db 6 GCCCTG 1
RESULT 6
US-09-360-237-4/c
; Sequence 4, Application US/09360237
; Patent No. 6322962
; GENERAL INFORMATION:
; APPLICANT: BROWN, MICHAEL S.
; APPLICANT: CHENG, DONG
; APPLICANT: ESPENSHADE, PETER J.
; APPLICANT: GOLDSTEIN, JOSEPH L.
; APPLICANT: RAWSON, ROBERT B.
; APPLICANT: SAKAI, JURO
; TITLE OF INVENTION: STEROL-REGULATED SITE-1 PROTEASE AND ASSAYS OF
; FILE REFERENCE: UTXD:567
; CURRENT APPLICATION NUMBER: US/09/360,237
; CURRENT FILING DATE: 1999-07-23
; EARLIER APPLICATION NUMBER: 60/096,571
; EARLIER FILING DATE: 1998-08-14
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 4338
; TYPE: DNA
; ORGANISM: Human
US-09-360-237-4

Query Match 12.7%; Score 124.4; DB 3; Length 4338;
Best Local Similarity 99.2%; Pred. No. 6.3e-23;
Matches 125; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTCGCGGCGGAACAGCGCTGGGCACTCCATTCCGGGCGCTGTTTACTCCCAACTCTCGCGAG 60
Db 126 CTCGCGGCGGAACAGCGCTGGGCACTCCATTCCGGGCGCTGTTTACTCCCAACTCTCGCGAG 67
QY 61 ACTGGGCGCGGCGGCGGCGAGGCCACACAGCTGGAGGCTCAGCTCCGCGGACCCAGCGT 120
Db 66 ACTGGGCGGCGGCGGCGGCGAGGCCACACAGCTGGAGGCTCAGCTCCGCGGACCCAGCGT 7

QY 121 GCCCTG 126
Db 6 GCCCTG 1

RESULT 7
US-09-891-711-3/c
; Sequence 3, Application US/09891711
; Patent No. 6723553
; GENERAL INFORMATION:
; APPLICANT: Pharmacia AB
; TITLE OF INVENTION: Promoter Sequences
; FILE REFERENCE: 00130
; CURRENT APPLICATION NUMBER: US/09/891,711
; CURRENT FILING DATE: 2001-06-26
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 4338
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (497)..(3655)
US-09-891-711-3

Query Match 12.7%; Score 124.4; DB 4; Length 4338;
Best Local Similarity 99.2%; Pred. No. 6.3e-23;
Matches 125; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTCGCGCGGAACAGCGCTGGGCACTCCATTCCGGGCTGTTTACTCCCAACTCTCGCGAG 60
Db |||||
126 CTCGCGCGGAACAGCGCTGGGCACTCCATTCCGGGCTGTTTACTCCCAACTCTCGCGAG 67
QY 61 ACTGGCGCGCGGCGAGCGAGCCACAGCTGGGAGCCTCAGCTCCGCGGACCCAGCGT 120
Db |||||
66 ACTGGCGGACCGGCGGCGAGCGAGCCACAGCTGGGAGCCTCAGCTCCGCGGACCCAGCGT 7
QY 121 GCCCTG 126
Db 6 GCCCTG 1
RESULT 8
US-09-949-016-542/c
; Sequence 542, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 542
; LENGTH: 4338
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-542

Query Match 12.7%; Score 124.4; DB 4; Length 4338;
Best Local Similarity 99.2%; Pred. No. 6.3e-23;
Matches 125; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CTCGCGCGGAACAGCGCTGGGCACTCCATTCCGGGCTGTTTACTCCCAACTCTCGCGAG 60
Db |||||
126 CTCGCGCGGAACAGCGCTGGGCACTCCATTCCGGGCTGTTTACTCCCAACTCTCGCGAG 67
QY 61 ACTGGCGCGCGGCGAGCGAGCCACAGCTGGGAGCCTCAGCTCCGCGGACCCAGCGT 120
Db |||||
66 ACTGGCGGACCGGCGGCGAGCGAGCCACAGCTGGGAGCCTCAGCTCCGCGGACCCAGCGT 7
QY 121 GCCCTG 126
Db 6 GCCCTG 1

RESULT 9
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765

; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2
Query Match 5.4%; Score 53; DB 3; Length 4403765;
Best Local Similarity 48.4%; Pred. No. 0.0074;
Matches 140; Conservative 2; Mismatches 147; Indels 0; Gaps 0;
QY 48 CAACTCTCCGAGACTGGGCGCGCGGCGAGAGCCACAGCTGGGAGCCTCAAGTCC 107
Db 1634333 CACCCGCGACCGGCGAGCGCGCGCGGATCTGTGGGCAACGCTGGGGCGG 1634274
QY 108 GCCGACCCAGCGTGCCTGTCTGCCGGCTCCCGGGGCTTCCGTCGCGGCTCTGACG 167
Db 1634273 GCGGCTCGGCGCGCGCCGCGCCAAAGTCGCGGGGCGCGGGGCTGTTCGCGAC 1634214
QY 168 CCGTGGGCGAGCGGACCGCGCGGAGGATGACGAAGGTGCTCGCGACATTTGCCGCGG 227
Db 1634213 CGGCGGCGCGGTGGGGCGGTGGGGCGCGCGCGCGGCGCGGGGTAGCGCGG 1634154
QY 228 CGGGGGCGGTGGCGAGGGTGGAAAGCGAGGGCGGTGGCGAGCGAGCTGCCAGCGCGGAG 287
Db 1634153 CTGGCTGTGGCAATGTTGGAGTCGGMGGGCGCGGCGGAGAGCTCTGCTGGYGGGC 1634094
QY 288 AACGCGCTGGGGGAACCCCTTGGTCCGCTCTGCGGCTGCTAGGATCC 336
Db 1634093 AACCGCGGGCGCGCGCGCAACCGCGGACTGTTTCGGGGTTCGCGGAAACC 1634045

RESULT 10
US-09-949-016-13632
; Sequence 13632, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13632
; LENGTH: 300402
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(300402)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13632

Query Match 5.3%; Score 52.2; DB 4; Length 300402;
Best Local Similarity 48.8%; Pred. No. 0.0044;
Matches 141; Conservative 0; Mismatches 148; Indels 0; Gaps 0;
QY 151 CGTGGCGCTCTGGAGCGCGTGGGCGAGCGGAGCACCGGAGGATGGACCAAGTGCT 210
Db |||||
2216 CGGCGCGGAGGTGCGGAGCGGAGCTCGCGCTCGCGCGGACTGGGGACCCGGGAGCC 2275
QY 211 CGCGACATTTGCGCGCGCGGGCGGTGGAGAGGTGGAAGCGAGGGGGGTGGCCAGCG 270
Db |||||
2276 TCGGGGTGCGCGCTGCAGAGTCCGGGAGGGGCGGGGAGCGCGGGCGCTGCACCGCG 2335

QY 271 AGCTGCCAGCGCGAGAACCGCTGGGGAAACCTTGGTCCGTCTGCGCGTCCGCTCTA 330
Db 2336 CGCGCTCGAGGTGGACGCGCGGTAGCCCGAGGGACAGCAGAGGGATCGCGC 2395
QY 331 GGATCCCGAAAGAGACAGCGGCGGAAAGGGCCAGGCTGGGCCAGGATCTAGAAGA 390
Db 2396 GGATCCCGCGCGGAGGGGTATCGGGTCCGGAACGCGCGCGGGGCACGCGGCGGAGCGG 2455
QY 391 CTGCTGGCGAGGCTCCCTGCCCCCGCGGCTGTCTCATGGAAGTGG 439
Db 2456 GGCCCCGGCGCGGCTGATGCTGCCGAGCCCGGCTCGCTTGGAGGGG 2504

RESULT 11
US-09-949-016-15924/c
; Sequence 15924, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15924
; LENGTH: 42053
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(42053)
; OTHER INFORMATION: n = A, T, C or G
US-09-949-016-15924

Query Match 4.9%; Score 48; DB 4; Length 42053;
Best Local Similarity 44.9%; Pred. No. 0.028;
Matches 183; Conservative 0; Mismatches 225; Indels 0; Gaps 0;

QY 21 GGCATCCATTCCGGGCTGTTTACTCCGAACTCTCGGAGACTGGCGGCGCGGCGAGG 80
Db 1181 GGCCTTTCGCGAGCGCGGATTCCTCCAACTTCTCCCTCGGCTCCCGCGCGGGGA 1122
QY 81 AGGCCACAGCTGGGAGCTCAGCTCCGCGGACCCAGCGTGCCTGTCTGTCCCGGCTC 140
Db 1121 AGAGGGACAGCGCGGGCCCCGAGAGGCTCGGCGCCCGCGCACCCGACCC 1062
QY 141 CGGGGCTTTCGTCGCGCTCTGAGCGCGTGGGAGCGGACACCGCGGAGGATGA 200
Db 1061 CTCGGCTTTCGCGGCGCGGCGGAGCGGCCCTCAGCGGGAGCGCGCCAGGAGGAAGAA 1002
QY 201 CGAAGTGTCTCGACATTTTCGCGCGCGCGGCGCGGCTGGCAGGTGGAGCGGAGGCG 260
Db 1001 AGAGGGCGAGGAGGAGAGGGGGAGCTTCGACAGAGAAAGCGAGGAGGAGCGCGG 942
QY 261 GTGGCCAGCGAGTGCAGCGCGGAGAAACGCGCTGGGGGAAACCTTGTTCGCTCTGG 320
Db 941 GCGGGCGGAGACCGCGGGGGCGCTTCAGACACGCGGCGAGTGCACAGCGCCCGCCGGA 882
QY 321 CGTCTGCTTAGGATCCCCGAAAGAGCAGCGGCGCGAAAGCGGCCAGGCTGGCGCAGA 380
Db 881 CCCCAGCGCGGAGCGCGGAGAGAAACGCGGGGCTGCGGGGAGAAAGCGGGGTCGGGG 822
QY 381 TCTAGAAAGACTGCTGGCGAGGCTCCCTGCCCCCGCGGCTGCTG 428

Db 821 GGAGAAACGGGGCGCGGAGGAGCGCGGAGCGCGGGGGAGAGCCG 774

RESULT 12
US-09-827-688-8
; Sequence 8, Application US/09827688
; Patent No. 6821955
; GENERAL INFORMATION:
; APPLICANT: ORSON, FRANK
; APPLICANT: KINSEY, BERMA
; APPLICANT: BHOGAL, BALBIR
; TITLE OF INVENTION: MACROAGGREGATED PROTEIN CONJUGATES AS ORAL GENETIC IMMUNIZATION
; TITLE OF INVENTION: AGENTS
; FILE REFERENCE: P01949US1/10004014
; CURRENT APPLICATION NUMBER: US/09/827,688
; CURRENT FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: 60/195,680
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 154746
; TYPE: DNA
; ORGANISM: HERPESVIRUS 2
US-09-827-688-8

Query Match 4.8%; Score 47.4; DB 4; Length 154746;
Best Local Similarity 49.8%; Pred. No. 0.065;
Matches 120; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

QY 60 GACTGGGCGCGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 119
Db 133282 GCGCGGCGGAGCGGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 133341
QY 120 TGCCCTGTCTCTCCCGCGCTCCCGGCGCTTCCGTCGCGCTCTGACCGCCGCTGGGACG 179
Db 133342 GGACGGCGCGGGGGAGCGGCGCGGGGGAGCGGCGCGGGGGCGCGGGGGCGG 133401
QY 180 GGACCGACCGCGGAGGATGAGCGAAGTGTCTCGGACATTTTCGCGCGCGGCGGCGGCGGTG 239
Db 133402 GGGGCGCGGGGGAGCGGCGGAGCGGCGGAGCGGCGGAGCGGCGGAGCGGCGGAGCGG 133461
QY 240 GCAGGTGGAAGCGGAGCGGCGTGCACGAGCGCTGCCAGCGCGGAGAAACGCGCTGGG 299
Db 133462 GACGCGGGGAGCGGCGGAGCGGCGGAGCGGCGGAGCGGCGGAGCGGCGGAGCGG 133521
QY 300 G 300
Db 133522 G 133522

RESULT 13
US-09-827-688-8/c
; Sequence 8, Application US/09827688
; Patent No. 6821955
; GENERAL INFORMATION:
; APPLICANT: ORSON, FRANK
; APPLICANT: KINSEY, BERMA
; APPLICANT: BHOGAL, BALBIR
; TITLE OF INVENTION: MACROAGGREGATED PROTEIN CONJUGATES AS ORAL GENETIC IMMUNIZATION
; TITLE OF INVENTION: AGENTS
; FILE REFERENCE: P01949US1/10004014
; CURRENT APPLICATION NUMBER: US/09/827,688
; CURRENT FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: 60/195,680
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 154746
; TYPE: DNA
; ORGANISM: HERPESVIRUS 2
US-09-827-688-8

Query Match 4.8%; Score 47.4; DB 4; Length 154746;
Best Local Similarity 49.8%; Pred. No. 0.065;
Matches 120; Conservative 0; Mismatches 121; Indels 0; Gaps 0;
QY 60 GACTGGCGCGCGGACAGGAGCCACAGCTGGAGGCTCAGCTCCGCCACCCAGCG 119
Db 148460 GCGCGGGGGACCGGCGGGGGGACGCGCGGGGGGACGCGCGGGGGGACGCGCGGG 148401
QY 120 TGCCCTGTCTGTCGCCGCTCCCGGGCTTGGTGGCGCTCTGACGCGCTGGGCGAG 179
Db 148400 GACGGCGCGGGGACGCGCGGGGGGACGCGCGGGGGGCGCGGGGGCG 148341
QY 180 GGACCAACCGCGGAGATGACGAAGTGTCTCCGACATTTTGGCGCGCGGGGGCGGTG 239
Db 148340 GCGGGCGGGGGACGCGGGGGGACGCGGGGGGACGCGGGGGGACGCGGGGG 148281
QY 240 CGAGGTGGAACGAGGCGCTGGCCAGCAGCTGCCAGGCGGCGAGAACCGCTGGGG 299
Db 148280 GACGGGGGACGCGGGGGGACGCGGGGGGACGCGGGGGGACGCGGGGGGACG 148221
QY 300 G 300
Db 148220 G 148220

RESULT 14
US-09-984-880-1/c
; Sequence 1, Application US/09984880
; Patent No. 6489153
; GENERAL INFORMATION:
; APPLICANT: Ming-Hui WEI
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; FILE REFERENCE: CL001305
; CURRENT FILING DATE: 2001-10-31
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2571
; TYPE: DNA
; ORGANISM: HomoSapien
US-09-984-880-1

Query Match 4.8%; Score 46.8; DB 4; Length 2571;
Best Local Similarity 49.7%; Pred. No. 0.02;
Matches 146; Conservative 0; Mismatches 147; Indels 1; Gaps 1;
QY 62 CTGGCGCGCGGCGGACGAGGCCACAGCTGGAGCCTCAGCTCCGCCACCCAGCGTG 121
Db 381 CAGAGCAGCCTGAGCAGCTGGCACCGCTGGAAGGGGCGCGCGGAGCTGGTGACGAGG 322
QY 122 CCTGTCTGTCCCGGCTCCCGGGGCTTGGTGGCGCTCTGAGCGCGCTGGGCGAGCGGG 181
Db 321 CGTGTGTGACGCGCGCGCGCGGACCGCGGGCGCGCGAGCGCGCTCGGGGTACGGGC 262
QY 182 AC-CACGCGGAGGATGACGAAGTGTCTCGCGACATTTGCGGCGCGGGGGCGGTGG 240
Db 261 ACGCACAGCAGTAGTGTGGCTCCCGGGGCGCGCGCGCGCGCGCGCGCGCGGGT 202
QY 241 CAGGGTGGAGCGGAGGCGGTGGCCAGCAGCTGCCAGGCGCGGAGAACCGCTGGGG 300
Db 201 GCGTCTGCGTCCCGGGGCGGTGGCGCGCTAGGGGCGGAGTGAAGCGAGTGCAGTCGGGA 142
QY 301 AACCTTGTGCTCGCTTGGCGGTCTAGGATCCCCGAAAAGGAGCACGGGC 354
Db 141 AGCTCAGGACGAAGCGGACGCGGGGAGCCATGGCCCCCAGCGAGACCCCGCGC 88

RESULT 15
US-10-277-032-1/c

; Sequence 1, Application US/10277032
; Patent No. 6664087
; GENERAL INFORMATION:
; APPLICANT: Ming-Hui WEI
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; FILE REFERENCE: CL001305 DIV
; CURRENT FILING DATE: 2002-10-22
; PRIOR FILING DATE: 2001-10-31
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2571
; TYPE: DNA
; ORGANISM: HomoSapien
US-10-277-032-1

Query Match 4.8%; Score 46.8; DB 4; Length 2571;
Best Local Similarity 49.7%; Pred. No. 0.02;
Matches 146; Conservative 0; Mismatches 147; Indels 1; Gaps 1;
QY 62 CTGGCGCGCGGCGGACGAGGCCACAGCTGGAGCCTCAGCTCCGCCACCCAGCGTG 121
Db 381 CAGAGCAGCCTGAGCAGCTGGCACCGCTGGAAGGGGCGCGCGGAGCTGGTGACGAGG 322
QY 122 CCTGTCTGTCCCGGCTCCCGGGGCTTGGTGGCGCTCTGAGCGCGCTGGGCGAGCGGG 181
Db 321 CGTGTGTGACGCGCGCGCGCGGACCGCGGGGCGCGCGAGCGCGCTCGGGGTACGGGC 262
QY 182 AC-CACGCGGAGGATGACGAAGTGTCTCGCGACATTTGCGGCGCGGGCGCGGTGG 240
Db 261 ACGCACAGGAGTAGTGTGGCTCCCGGGGCGCGCGCGCGCGCGCGCGCGGGT 202
QY 241 CAGGGTGGAGCGGAGGCGGTGGCCAGCAGCTGCCAGGCGCGGAGAACCGCTGGGG 300
Db 201 GCGTCTGCGTCCCGGGGCGGTGGCGCGCTAGGGGCGGAGTGAAGCGAGTGCAGTCGGGA 142
QY 301 AACCTTGTGCTCGCTTGGCGGTCTAGGATCCCCGAAAAGGAGCACGGGC 354
Db 141 AGCTCAGGACGAAGCGGACGCGGGGAGCCATGGCCCCCAGCGAGACCCCGCGC 88

Search completed: October 9, 2005, 03:28:30
Job time : 220 secs

This Page Blank (uspiv,

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2005, 00:48:23 ; Search time 3169 Seconds
(without alignments)
11771.216 Million cell updates/sec

Title: US-10-829-118-1

Perfect score: 980

Sequence: 1 ctccgcggcgaacacgcctg.....ctgcaccacgtaagtctga 980

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1:*

2: gb_est2:*

3: gb_hic:*

4: gb_est3:*

5: gb_est4:*

6: gb_est5:*

7: gb_est6:*

8: gb_gss1:*

9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	471.6	48.1	500	6	CB049709
C 2	468	47.8	468	1	AI792242
C 3	451.8	46.1	455	1	AI760651
C 4	448	45.7	450	2	AW051661
C 5	440.6	45.0	1240	5	BQ437878
C 6	435	44.4	435	1	AI685715
C 7	435	44.4	435	1	AI733521
C 8	435	44.4	435	2	BF446188
C 9	432	44.1	432	1	AI554846
C 10	430.2	43.9	436	2	AW005919
C 11	411.8	42.0	415	1	AI362974
C 12	404	41.2	404	2	BE048839
C 13	362.8	37.0	389	1	AI498080
C 14	303.6	31.0	457	2	BF570483
C 15	263	26.8	263	1	AI198389
C 16	249	25.4	251	5	BX100137
C 17	163	16.6	582	5	BP364386
C 18	159.6	16.3	567	7	CR537300
C 19	141.8	14.5	584	5	BP358100
C 20	137.6	14.0	742	7	CR790685
C 21	133.6	13.6	649	7	CR750317
C 22	130.2	13.3	390	7	CR549288
C 23	130.2	13.3	536	7	CR549308
C 24	128.8	13.1	575	1	AL701753

C 25	127.2	13.0	582	5	BP357251
C 26	125	12.8	701	6	CD617566
C 27	123.4	12.6	249	4	BM146635
C 28	121.8	12.4	564	5	AX957183
C 29	120.2	12.3	144	1	AA998897
C 30	119.2	12.2	350	7	CR545379
C 31	116.4	11.9	990	5	BX443626
C 32	112.4	11.5	506	4	BM194303
C 33	112.2	11.4	501	7	CR556966
C 34	110.2	11.2	829	4	BM553744
C 35	105.8	10.8	257	5	BU431331
C 36	105.8	10.8	505	5	BU431330
C 37	104.6	10.7	569	1	AL701734
C 38	92	9.4	596	4	BI545503
C 39	90.4	9.2	780	4	BG702439
C 40	90.2	9.2	454	1	AA045615
C 41	82.6	8.4	932	9	CNS00720
C 42	75.2	7.7	1300	9	AG430756
C 43	74	7.6	581	5	BP222057
C 44	73.4	7.5	879	9	AG130935
C 45	72.2	7.4	1165	9	AG030649

ALIGNMENTS

RESULT 1
CB049709/c
LOCUS NISC_gj12f01.xl NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:3271441
DEFINITION 3', mRNA sequence.
ACCESSION CB049709
VERSION EST.
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 500)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
COMMENT National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
CDNA Library Preparation:
DNA Sequencing Arrayed by: The I.M.A.G.E. Consortium/LLNL
Sequencing Center (NISC)
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
info@image.llnl.gov
Plate: LLAM8008 row: L column: 2
Seq primer: -21M13 forward primer (ABI).
Location/Qualifiers
1. 500
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3271441"
/sex="male"
/dev_stage="adult"
/clone_host="DH10B"
/clone_lib="NCI_CGAP_Pr28"
/note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
with a modified polylinker; plasmid DNA from the
normalized library NCI CGAP Pr22 was prepared, and ss
circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (clonoids
985608-986759, 1101192-1101959, and 1217928-1220615).
Subtraction by Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 48.1%; Score 471.6; DB 6; Length 500;
Best Local Similarity 99.2%; Pred. No. 2.9e-105;
Matches 474; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 300 GAACCTTGGTCCGCTCTCGCGTCTCTAGATCCCGAAAGAGACGCGGCGGAA 359
DB 495 GAGGCTTGGTCCGCTCTCGCGTCTCTAGATCCCGAAAGAGACGCGGCGGAA 436
QY 360 AGGGCAGGCTGGGCGAGGCTAGAAAGCTGCTGGCGAGGCTCCCTCCCGCGG 419
DB 435 AGCGGCAGGCTGGGCGAGGCTAGAAAGCTGCTGGCGAGGCTCCCTCCCGCGG 376
QY 420 GGCCTCTCTGATGAGTCTGCTGAGAGCTCGCTTCCCGCGGACCTTCTTCCGCGGGT 479
DB 375 GGCCTCTCTGATGAGTCTGCTGAGAGCTCGCTTCCCGCGGACCTTCTTCCGCGGGT 316
QY 480 CCACTGTCAGGACACCGCGGCTCGGACACCCACCCCGCGGCGGACCTGCGCTGGTG 539
DB 315 CCACGTCAGGACACCGCGGCTCGGACACCCACCCCGCGGCGGACCTGCGCTGGTG 256
QY 540 CCCCCTTAAACCCGGCGGTAGCTGTTAAGATGCGGAGTGTCCGTCGCGAAACGCGAA 599
DB 255 CCCCCTTAAACCCGGCGGTAGCTGTTAAGATGCGGAGTGTCCGTCGCGAAACGCGAA 196
QY 600 ACCCAAAATCCCGCTTCCCGGACCTCTGACCCCGCGGCGGACGACAGACTGGGC 659
DB 195 ACCCAAAATCCCGCTTCCCGGACCTCTGACCCCGCGGCGGACGACAGACTGGGC 136
QY 660 CTCGCCAGCGGACGCGCTGCGCGGACACCGGTGCGTGGCGAAACGCGAGACTTTGTAA 719
DB 135 CTCGCCAGCGGACGCGCTGCGCGGACACCGGTGCGTGGCGAAACGCGAGACTTTGTAA 76
QY 720 CGCCAGCTGTTGCTCTTTTGAAGAAACAGAAATGTTTAACTGTCGAAA 777
DB 75 CGCCAGCTGTTGCTCTTTTGAAGAAACAGAAATGTTTAACTGTCGAAA 18

RESULT 2
AI792242
LOCUS
DEFINITION
ov05e09.y5 NCI_CGAP_Kid3 Homo sapiens cDNA clone IMAGE:1636456 5',
mRNA sequence.

ACCESSION
AI792242
VERSION
AI792242.1 GI:5339949
KEYWORDS
EST.
SOURCE
Homo sapiens (human)

ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 468)
NCI/NIDR-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute / National Institute of Dental Research,
Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)
Other ESTs: ov05e09.x5

COMMENT
Contact: Robert Strausberg, ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

This read is a RESEQUENCE of a previously sequenced human clone
Original clone citation: see original entry for original citation
information
This 5' resequenced clone has no previous 5' data to verify this
new read against

FEATURES

Source

Putative full length read
The vector to vector length is 469
Seq primer: -40RP from Gibco.

Location/Qualifiers
1..468
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1636456"
/lab_host="DH10B"
/clone_lib="NCI CGAP Kid3"
/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer,
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT7T3 vector. mRNA
source: 2 pooled kidneys. Library went through one round
of normalization. Library constructed by Bento Soares and
M. Fatima Bonaldo."

ORIGIN

Query Match 47.8%; Score 468; DB 1; Length 468;
Best Local Similarity 100.0%; Pred. No. 2.2e-104; Indels 0; Gaps 0;
Matches 468; Conservative 0; Mismatches 0;

QY 304 CTTTGGTCCGCTCTCGCGTCTCTAGGATCCCCGAAAGAGACGCGGCGGAAAGCG 363
DB 1 CTTTGGTCCGCTCTCGCGTCTCTAGGATCCCCGAAAGAGACGCGGCGGAAAGCG 60
QY 364 GCCAGCTGGGCGAGGATCTAGAAAGACTGCTTCCCGCGGAGGCTCCTTGCCTCCCGCGGGCC 423
DB 61 GCCAGCTGGGCGAGGATCTAGAAAGACTGCTTCCCGCGGAGGCTCCTTGCCTCCCGCGGGCC 120
QY 424 TGCTGTGATGAGTCTGTGAGAGTCTGCTTCCCGCGGAGGCTTCCCTGTCAGGGTCCAC 483
DB 121 TGCTGTGATGAGTCTGTGAGAGTCTGCTTCCCGCGGAGGCTTCCCTGTCAGGGTCCAC 180
QY 484 GTCCAGGACACCGCGGCTCGGACACCCACCCCGCGGCGGACCTGCTGGGTGCCCC 543
DB 181 GTCCAGGACACCGCGGCTCGGACACCCACCCCGCGGCGGACCTGCTGGGTGCCCC 240
QY 544 TTAAACCCCGGCGGTAGTCTGTTAAGATGGCGAAGTGTGCGTCCGGAACACGCGAAACCC 603
DB 241 TTAAACCCCGGCGGTAGTCTGTTAAGATGGCGAAGTGTGCGTCCGGAACACGCGAAACCC 300
QY 604 CAAATCCCGCTGCCGACCTCTGACCCCGCGGCGGCGGACGACAGACTGGGCTCC 663
DB 301 CAAATCCCGCTGCCGACCTCTGACCCCGCGGCGGCGGACGACAGACTGGGCTCC 360
QY 664 CGACGCGGACGCGCTGCGGCGGACACCGGTGCGTGGCGAAACGCGAGGACCTTTGTAACGCC 723
DB 361 CGACGCGGACGCGCTGCGGCGGACACCGGTGCGTGGCGAAACGCGAGGACCTTTGTAACGCC 420
QY 724 ACGTGTGCTCTTTTGAAGAAACAGAAATGTTTAACTGTC 771
DB 421 ACGTGTGCTCTTTTGAAGAAACAGAAATGTTTAACTGTC 468

RESULT 3

AI760651/c

LOCUS

DEFINITION

wi66g09.x1 NCI_CGAP_Kid12 Homo sapiens cDNA clone IMAGE:2398336 3',
mRNA sequence.

ACCESSION

AI760651

VERSION

AI760651.1 GI:5176318

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 455)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 567 Std Error: 0.00
Seq primer: -40UP from Gibco.
FEATURES
source
1..455
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2398336"
/tissue_type="2 pooled tumors (clear cell type)"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Kid12"
/notes="Organ: Kidney; Vector: p7T73D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not I; Site 2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP_Kid5 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneIDs 1323912-1325831, 1471368-1472903 and
1492104-1493255). Subtraction by Bento Soares and M.
Fatima Bonaldo."

ORIGIN

Query Match 46.1%; Score 451.8; DB 1; Length 455;
Best Local Similarity 99.6%; Pred. No. 2.2e-100;
Matches 453; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 323 TCCTCTAGATCCCGAAAGAGACACGGGCGGAAAGGGCGGAGCTGGGCGAGGATC 382
DB 455 TCCTCTAGATCCCGAAAGAGACACGGGCGGAAAGGGCGGAGCTGGGCGAGGATC 396
QY 383 TAGAAGACTGCTGCGGAGGCTCCCTGCCCCCGGGGCTGCTGTCATGGACTCGTGG 442
DB 395 TAGAAGACTGCTGCGGAGGCTCCCTGCCCCCGGGGCTGCTGTCATGGACTCGTGG 336
QY 443 AGAGCTCGCTTCCCGCGCGGACCTTCTGCGAGGGTCCACGTCAGGACCGCGGCTC 502
DB 335 AGAGCTCGCTTCCCGCGCGGACCTTCTGCGAGGGTCCACGTCAGGACCGCGGCTC 276
QY 503 GGACACCCACCCCGCGCGGCACTGCGCTGGGTGCCCCCTTAAACCGGGCGGTAGCTC 562
DB 275 GGACACCCACCCCGCGCGGCACTGCGCTGGGTGCCCCCTTAAACCGGGCGGTAGCTC 216
QY 563 GTTAAGATGCGGAAGTGCCTGCGTCCGGAACACCGCAACCCCAATCCCGCTGCCGAC 622
DB 215 GTTAAGATGCGGAAGTGCCTGCGTCCGGAACACCGCAACCCCAATCCCGCTGCCGAC 156
QY 623 CTCCTGACCCCGCGCGGACGACAGCTGGGCTCCCGACGCGAGCGCTGCC 682
DB 155 CTCCTGACCCCGCGCGGACGACAGCTGGGCTCCCGACGCGAGCGGCTGCC 96
QY 683 GGACACCGGTGCGTCCGAAACGGAGGACTTTGTAAACGCAAGTGTGCTCTTTTGA 742
DB 95 GGACACCGGTGCGTCCGAAACGGAGGACCTTTGTAAACGCAAGTGTGCTCTTTTGA 36
QY 743 AAAACAAGATAAATGCTTAACCTGCTGAAA 777
DB 35 AAAACAAGATAAATGCTTAACCTGCTGAAA 1

RESULT 4
AW051661/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AW051661
wx27b06.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2544851 3',
mRNA sequence.
AW051661
AW051661.1 GI:5913892
Homo sapiens
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 450)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 540 Std Error: 0.00
Seq primer: -40UP from Gibco.
FEATURES
source
1..450
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2544851"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Kid11"
/notes="Organ: Kidney; Vector: p7T73D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not I; Site 2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneIDs 1322376-1323911, 1456007-1456775, and
1500552-1502855). Subtraction by Bento Soares and M.
Fatima Bonaldo."

ORIGIN

Query Match 45.7%; Score 448; DB 2; Length 450;
Best Local Similarity 100.0%; Pred. No. 1.9e-99;
Matches 448; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 324 CGCTCTAGGATCCCGAAAGAGACACGGGCGCGAAAGCGGCGGAGGATCT 383
DB 450 CGCTCTAGGATCCCGAAAGAGACACGGGCGCGAAAGCGGCGGAGGATCT 391
QY 384 AGAAGACTGCTGCGGAGGCTCCCTGCCCCCGGGGCTGCTGTCATGGACTCGTGA 443
DB 390 AGAAGACTGCTGCGGAGGCTCCCTGCCCCCGGGGCTGCTGTCATGGACTCGTGA 331
QY 444 GAGCTCGCTTCCCGCGCGGACCTTCTGCGAGGGTCCACGTCAGGACCGCGGCTCG 503
DB 330 GAGCTCGCTTCCCGCGCGGACCTTCTGCGAGGGTCCACGTCAGGACCGCGGCTCG 271
QY 504 GACACCCACCCCGCGCGGACCTGCGCTGGGTGCCCTTTAAACCGGGCGGTAGCTCG 563
DB 270 GACACCCACCCCGCGCGGACCTGCGCTGGGTGCCCTTTAAACCGGGCGGTAGCTCG 211
QY 564 TTAAGATGCGAAGTGTCCGGTCCGGAACACCGGAAACCCCAATCCCGCTGCCGACC 623
DB 210 TTAAGATGCGAAGTGTCCGGTCCGGAACACCGGAAACCCCAATCCCGCTGCCGACC 151

```

624 TCCTGACCCCGGCGCCCGGACGACGACACTGGGCTCCCGACGGGCGCGCTGCGG 683
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
150 TCCTGACCCCGGCGCCCGGACGACGACACTGGGCTCCCGACGGGCGCGCTGCGG 91
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
684 GACACCGGCTGGTGGAAACGGAGGACCTTTGTAAACGCCAGTGTGCTCTTTTGA 743
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
90 GACACCGGCTGGTGGAAACGGAGGACCTTTGTAAACGCCAGTGTGCTCTTTTGA 31
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
744 AAAACAAGATAAATGTGTTAAATGTC 771
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
30 AAAACAGATAAATGTGTTAAATGTC 3
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 5
BQ437878
LOCUS
DEFINITION BQ437878 1240 bp mRNA linear EST 24-MAY-2002
AGENCOURT_7890066 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:6176570
5', mRNA sequence.
ACCESSION BQ437878
VERSION BQ437878.1 GI:21176954
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE NIH-MGC http://mgs.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1353 row: c column: 03
High quality sequence stop: 421.
Location/Qualifiers
1. .1240
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6176570"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: uterus; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2.1 kb."

FEATURES
source
1. 45.0%; Score 440.6; DB 5; Length 1240;
Best Local Similarity 94.4%; Pred. No. 1.4e-97;
Matches 490; Conservative 0; Mismatches 24; Indels 5; Gaps 3;

274 TGCACGGCGGAGAACGGCTGGGGAAACCTTGTTCGGCTTGGCGTCTAGGA 333
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1 TGCACGGCGGAGAACGGCTGGGGAAACCTTGTTCGGCTTGGCGTCTAGGA 60
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

334 TCCCGGAAAAGGAGCAGCGGCGGAAAGCGGCAGGCTGGGCCAGGATCTAGAAAGCTG 393
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
61 TCCCGGAAAAGGAGCAGCGGCGGAAAGCGGCAGGCTGGGCCAGGATCTAGAAAGCTG 120
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

394 CTGTGGCGCAGGCTCCCTGCCCCCGCGGCGCTGTGTATGAGTCTGTGAGAGCTCGCTT 453
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
121 CTGTGGCGCAGGCTCCCTGCCCCCGCGGCGCTGTGTATGAGTCTGTGAGAGCTCGCTT 180
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

454 CCGCGCGGACCCCTTCTGTGAGGGTCCAGTCCAGGCACCGGGCTGGACACCCAC 513
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
181 CCGCGCGGACCCCTTCTGTGAGGGTCCAGTCCAGGCACCGGGCTGGACACCCAC 240
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

514 CCCGGCGGGGCACTGCCCTGGGTGCCCTTAACCCGGGCGGTAGCTCGTTAAGATGCG 573
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
241 CCCGGCGGGGCACTGCCCTGGGTGCCCTTAACCCGGGCGGTAGCTCGTTAAGATGCG 300
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

574 GAAGTGTCCGCTCGGAAACACGCAAAATCCCAATCCCGCTGCCGCGACCTCCTGACCCC 633
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
301 GAAGTGTCCGCTCGGAAACACGCAAAATCCCAATCCCGCTGCCGCGACCTCCTGACCCC 360
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

634 CGGCCCGGACGACGACACTGGGCTTCCGACGCGCAGCGCTGCGGAGACCC-GG 692
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
361 CGGCCCGGACGACGACACTGGGCTTCCGACGCGCAGCGCTGCGGAGACCCGGG 420
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

693 TCGGTGCGAAACGG--AGGACCTTTGTAACGCCAGCTG--TTTGTCTTTTGAAGAAC 748
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
421 TCGGTGCGAAACGGCAGGACCCCTTTGTAAACGCCAGCTGTTGTCTTTTGAAGAAC 480
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

749 AAGATAAATGTGTTAAACTGTCTGAAAGCTTGCCGCC 787
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
481 CAGAATAAATGCGCTCAACTGCGCTGAAAGAGTCTGCC 519
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 6
AI685715/c
LOCUS
DEFINITION AI685715 435 bp mRNA linear EST 07-MAR-2000
tu37c07.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2253228 3',
mRNA sequence.
ACCESSION AI685715
VERSION AI685715.1 GI:4897009
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 435)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 562 Std Error: 0.00
Seq primer: -40UP from Gibco.
Location/Qualifiers
1. 435
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2253228"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Pr28"
/notes="Organ: prostate; Vector: pVT73D-Pac (Pharmacia)
with a modified polylinker; Plasmid DNA from the
normalized library NCI CGAP Pr22 was prepared, and as
circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (clonoids
985608-986759, 1101192-1101959, and 1217928-1220615).
Subtraction by Bento Soares and M. Fatima Bonaldo."

ORIGIN
Query Match 44.4%; Score 435; DB 1; Length 435;

```


Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
cDNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -40UP from Gibco.

FEATURES

Source

1. 435
/organism="Homo sapiens"
/mol_type="mrna"
/db_xref="taxon:9606"
/clone="IMAGE:3647745"
/sex="male"
/dev stage="adult"
/lab_host="DH10B"
/clone_lib="NCI-CGAP Pr28"
/note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
with a modified polylinker; Plasmid DNA from the
normalized library NCI CGAP Pr22 was prepared, and ss
circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (clonoids
985608-986759, 110192-1101959, and 1217928-1220615).
Subtraction by Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 44.4%; Score 435; DB 2; Length 435;
Best Local Similarity 100.0%; Pred. No. 3e-96; Indels 0; Gaps 0;
Matches 435; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 337 CGAAGAAGGACGACGGCGGAAAGCGGAGGCTGGCGCAGGATCTAGAAAGACTGCGCT 396
Db 435 CGAAGAAGGACGACGGCGGAAAGCGGAGGCTGGCGCAGGATCTAGAAAGACTGCGCT 376
QY 397 GCGCAGGCTCCCTGCGCCCGCGGGGCTGCTGTATGAGTCTGTGGAGAGCTGCTTCCC 456
Db 375 GCGCAGGCTCCCTGCGCCCGCGGGGCTGCTGTATGAGTCTGTGGAGAGCTGCTTCCC 316
QY 457 GCGCGAGCCTTCTCGAGGGGTTCAGCTCCAGGACACGGCGGCTGGACACCCACCC 516
Db 315 GCGCGAGCCTTCTCGAGGGGTTCAGCTCCAGGACACGGCGGCTGGACACCCACCC 256
QY 517 GCGCGGGCACCTGCGCTGGTGCCTTAAACCGGGCGGTAGCTCGTTAAGATGCGAA 576
Db 255 GCGCGGGCACCTGCGCTGGTGCCTTAAACCGGGCGGTAGCTCGTTAAGATGCGAA 196
QY 577 GTGTCCGGTCCGGAACACGCGAAACCCCAATCCCGCTGCGGACCTCTGACCCCGG 636
Db 195 GTGTCCGGTCCGGAACACGCGAAACCCCAATCCCGCTGCGGACCTCTGACCCCGG 136
QY 637 CCCCACGGGACGACAGACTGGGCTCCCGACGCGAGCGGCTGCGGGGACACCGGTGG 696
Db 135 CCCCACGGGACGACAGACTGGGCTCCCGACGCGAGCGGCTGCGGGGACACCGGTGG 76
QY 697 TCGGAACCGAGGACCTTTTGAACGCGACGTGTTGCTCTTTTGAAGAAACCAAGATAA 756
Db 75 TCGGAACCGAGGACCTTTTGAACGCGACGTGTTGCTCTTTTGAAGAAACCAAGATAA 16
QY 757 ATGTGTTAAACTGTC 771
Db 15 ATGTGTTAAACTGTC 1

RESULT 9

AI554846/c 432 bp mRNA linear EST 23-MAR-1999
LOCUS t945b04.x1 Soares_NhMPu_S1 Homo sapiens cDNA clone IMAGE:2089615
DEFINITION 3', mRNA sequence.
ACCESSION AI554846
VERSION AI554846.1 GI:4487209

KEYWORDS

SOURCE
ORGANISM

EST. Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

TITLE

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

1 (bases 1 to 432)

Unpublished (1997)

Journal

Comment

Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 430.

FEATURES

Source

1. 432
/organism="Homo sapiens"
/mol_type="mrna"
/db_xref="taxon:9606"
/clone="IMAGE:2089615"
/tissue type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/clone_lib="Soares_NhMPu_S1"
/note="Organ: mixed (see below); Vector: pT7T3D-Pac
(Pharmacia) with a modified polylinker; Site 1: Not I;
Site 2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NBHM, pregnant uterus
NBHPU, and fetal heart NBH19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."

ORIGIN

Query Match 44.1%; Score 432; DB 1; Length 432;
Best Local Similarity 100.0%; Pred. No. 1.6e-95;
Matches 432; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 340 AAAAGAGACACGGGCGGAAAGCGGCGGAGGCTGGCGCAGGATCTAGAAAGACTGCGTGGC 399
Db 432 AAAAGAGACACGGGCGGAAAGCGGCGGAGGCTGGCGCAGGATCTAGAAAGACTGCGTGGC 373
QY 400 GCAGGCTCCCTGCGCCCGCGGGGCTGCTGTATGAGTCTGTGGAGAGCTGCTTCCCGG 459
Db 372 GCAGGCTCCCTGCGCCCGCGGGGCTGCTGTATGAGTCTGTGGAGAGCTGCTTCCCGG 313
QY 460 CCGACCTTCTCTGAGGGGTCCACCTCCAGGACACCGCGGCTCGGACACCCACCCCGG 519
Db 312 CCGACCTTCTCTGAGGGGTCCACCTCCAGGACACCGCGGCTCGGACACCCACCCCGG 253
QY 520 CCGGCGACCTGCGCTCGGTGCGCCCTTAAACCGGGCGGTAGCTCGTTAAGATGCGGAAGTG 579
Db 252 CCGGCGACCTGCGCTCGGTGCGCCCTTAAACCGGGCGGTAGCTCGTTAAGATGCGGAAGTG 193
QY 580 TCGGGTCCGGAACACGCGAAACCCCAATCCCGCTGCGGACCTCTGACCCCGGCCC 639
Db 192 TCGGGTCCGGAACACGCGAAACCCCAATCCCGCTGCGGACCTCTGACCCCGGCCC 133
QY 640 CACGGACACACAGACTGGGCTCCCGACGCGAGCGGCTGCGGGACACCGGTGCGTGC 699
Db 132 CACGGACACACAGACTGGGCTCCCGACGCGAGCGGCTGCGGGACACCGGTGCGTGC 73
QY 700 GAAACGAGGACCTTTGTAAACGCGCAGTGTGCTCTTTTGAAGAAACCAAGATAAATG 759
Db 72 GAAACGAGGACCTTTGTAAACGCGCAGTGTGCTCTTTTGAAGAAACCAAGATAAATG 13
QY 760 TGTATAACTGTC 771
Db 15 TGTATAACTGTC 1

```

Db      12  TGTAAACTGTC 1

RESULT 10
AW005919/c
LOCUS   w290e09.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2566120 3',
DEFINITION mRNA sequence.
ACCESSION AW005919
VERSION   AW005919.1 GI:5854686
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens

REFERENCE
1 (bases 1 to 436)
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS  National Cancer Institute / National Institute of Neurological
TITLE     Disorders and Stroke, Brain Tumor Genome Anatomy Project
(JC GAP/ BT GAP), Tumor Gene Index
Unpublished (1998)
JOURNAL   Contact: Robert Strausberg, Ph.D.
COMMENT   Email: cgabbs-r@mail.nih.gov
          Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
          Ph.D.
          cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
          Bonaldo, Ph.D.
          DNA Sequencing by: Greg Lennon, Ph.D.
          Clone distribution: NCI-CGAP clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          www-bio.llnl.gov/bbrp/image/image.html
          Insert Length: 541 Std Error: 0.00
          Seq primer: -40UP from Gibco.

FEATURES             source
    1..436
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /clone="IMAGE:2566120"
        /tissue_type="anaplastic oligodendroglioma"
        /lab_host="DH10B"
        /clone_lib="NCI_CGAP_Brn25"
        /notes="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
        modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
        strand cDNA was primed with a Not I - oligo(dT) primer [5'
        TGTTACCAATCTGAAGTGGAGCGCGGCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
        T 3']; double-stranded cDNA was ligated to Eco RI
        adaptors (Pharmacia), digested with Not I and cloned into
        the Not I and Eco RI sites of the modified pT7T3 vector.
        Library is normalized, and was constructed by Bento
        Soares and M.Fatima Bonaldo."

ORIGIN
Query Match      43.9%; Score 430.2; DB 2; Length 436;
Best Local Similarity 99.3%; Pred. No. 4.5e-95;
Matches 432; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 343 AGAGACACGGGCGCGAAAGCGCGCAGCTGGCGCAGATCTAGAAAGCTGCTGCGCGCA 402
Db   |
QY 436 AGAGACACGGGCGCGAAAGCGCGCAGCTGGCGCAGATCTAGAAAGCTGCTGCGCGCA 377
Db   |
QY 403 GGTCTCTCCCGCGCGGCGCTGCTCATGACTCGTGAGAGCTCGTTCCTCCGCGCG 462
Db   |
QY 376 GGTCTCTCCCGCGCGGCGCTGCTCATGACTCGTGAGAGCTCGTTCCTCCGCGCG 317
Db   |
QY 463 ACCCTTCTCTCGAGGGGTCCACGCTCCAGGACACCGCGGCTCGGACACCCACCGCGCG 522
Db   |
QY 316 ACCCTTCTCTCGAGGGGTCCACGCTCCAGGACACCGCGGCTCGGACACCCACCGCGCG 257
Db   |
QY 523 GGCACCTGCGCTGGGTGCGCCCTTAACCCGGGGCGGTAGCTCGTTAAGATGCGGAAGTGTCC 582
Db   |
QY 256 GGCACCTGCGCTGGGTGCGCCCTTAACCCGGGGCGGTAGCTCGTTAAGATGCGGAAGTGTCC 197

```

363 GGCAGGCTGGCCAGGATCTAGAAAGACTGCTGGCGCAGGCTCCCTGGCCCCCGGGGC 422
|||||
Db |||||
415 GGCAGGCTGGCCAGGATCTAGAAAGACTGCTGGCGCAGGCTCCCTGGCCCCCGGGC 356
|||||
Qy |||||
423 CTGCTGTCATGACCTCGTGGAGAGCTCGCTTCCCGCGGAGCCCTTCCTGCAAGGGTCCA 482
|||||
Db |||||
355 CTGCTGTCATGACCTCGTGGAGAGCTCGCTTCCCGCGGAGCCCTTCCTGCAAGGGTCCA 296
|||||
Qy |||||
483 CGTCAGGACACCGGGGCTCGGACACCCACCCCGCGCGGCACTGCGCTTGGGTGCCC 542
|||||
Db |||||
295 CGTCAGGACACCGGGGCTCGGACACCCACCCCGCGCGGCACTGCGCTTGGGTGCCC 236
|||||
Qy |||||
543 CTTAAACCGGGGCGGTAGCTGCTTAAGATGGGAAAGTCCCGTCCGGAAACACGCGAAACC 602
|||||
Db |||||
235 CTTAAACCGGGGCGGTAGCTGCTTAAGATGGGAAAGTCCCGTCCGGAAACACGCGAAACC 176
|||||
Qy |||||
603 CCAATCCCGCTGCGGACCTCTGACCCCGCGGCGGACGAGGACGACAGCTGGGCTC 662
|||||
Db |||||
175 CCAATCCCGCTGCGGACCTCTGACCCCGCGGCGGACGAGGACGACAGCTGGGCTC 116
|||||
Qy |||||
663 CGAGCGCAGCGCTGCGGCGGACACCGGTGCTCGGAAACGAGGACCTTTGTAACGC 722
|||||
Db |||||
115 CGAGCGCAGCGCTGCGGCGGACACCGGTGCTCGGAAACGAGGACCTTTGTAACGC 56
|||||
Qy |||||
723 CACGTGTTGCTCTTTTGAAGAAACAGAAATAAATGTTTAACTGTGAAAA 777
|||||
Db |||||
55 CACGTGTTGCTCTTTTGAAGAAACAGAAATAAATGTTTAACTGTGAAAA 1

RESULT 12
LOCUS BE048839/c
DEFINITION hr54e04.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:3132318 3',
mRNA sequence.
ACCESSION BE048839
VERSION BE048839.1 GI:8365894
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 404)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -40UP from Gibco.
Location/Qualifiers
1..404
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3132318"
/lab_host="DH10B"
/clone_lib="NCI_CGAP Kid11"
/note="Organ: Kidney; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site: 1: Not 1; Site 2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP Kid3 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library

FEATURES
source

(cloneIds 132376-1323911, 1456007-1456775, and
1505552-1502855). Subtraction by Bento Soares and M.
Fatima Bonaldo. "

ORIGIN

Query Match 41.2%; Score 404; DB 2; Length 404;
Best Local Similarity 100.0%; Pred. No. 1.3e-88;
Matches 404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 370 CTGGCCAGGATCTAGAAAGACTGCTGGCGCAGGCTCCCTGGCCCCCGGGGCTGCTGT 429
|||||
Db 404 CTGGCCAGGATCTAGAAAGACTGCTGGCGCAGGCTCCCTGGCCCCCGGGGCTGCTGT 345
|||||
Qy 430 CATGACCTCGTGGAGAGCTCGCTTCCCGCGGAGCCCTTCCTGCAAGGGTCCAACCTCCAG 489
|||||
Db 344 CATGACCTCGTGGAGAGCTCGCTTCCCGCGGAGCCCTTCCTGCAAGGGTCCAACCTCCAG 285
|||||
Qy 490 GCACCGCGGCTCGGACACCCACCCCGCGCGGCACTGCGCTGGGTGCCCTTTAAACC 549
|||||
Db 284 GCACCGCGGCTCGGACACCCACCCCGCGCGGCACTGCGCTGGGTGCCCTTTAAACC 225
|||||
Qy 550 CGGCGGTAGCTGCTTAAGATGGGAAAGTGTCCGTCCGGAAACACGCGAAACCCCAATC 609
|||||
Db 224 CGGCGGTAGCTGCTTAAGATGGGAAAGTGTCCGTCCGGAAACACGCGAAACCCCAATC 165
|||||
Qy 610 CGGCTGCGGACCTCTGACCCCGCGGCGGACGAGACTGGGCTCCGACGC 669
|||||
Db 164 CGGCTGCGGACCTCTGACCCCGCGGCGGACGAGACTGGGCTCCGACGC 105
|||||
Qy 670 GCAGCGCTGCGGCGGACACCGGTGCTCGGAAACGAGGACCTTTGTAACGCCAGCTGT 729
|||||
Db 104 GCAGCGCTGCGGCGGACACCGGTGCTCGGAAACGAGGACCTTTGTAACGCCAGCTGT 45
|||||
Qy 730 TTGCTCTTTTGAAGAAACAGAAATAAATGTTTAACTGTCTG 773
|||||
Db 44 TTGCTCTTTTGAAGAAACAGAAATAAATGTTTAACTGTCTG 1

RESULT 13
LOCUS AI498080/c
DEFINITION tm66a04.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2163054 3',
mRNA sequence.
ACCESSION AI498080
VERSION AI498080.1 GI:4390062
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 389)
NCI/NINDS-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/STGAP), Tumor Gene Index
Unpublished (1998)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Insert Length: 516 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 378.
Location/Qualifiers
1..389
/organism="Homo sapiens"

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2163054"
/tissue_type="anaplastic oligodendroglioma"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Brn25"
/note="Organ: Brain; Vector: pTT3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGAGCGCGGCATAGTTTATTTTATTTTATTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pTT3 vector. Library is normalized, and was constructed by Bento Soares and M.Fatima Bernaldo."

ORIGIN

Query Match 37.0%; Score 362.8; DB 1; Length 389;
Best Local Similarity 99.2%; Pred. No. 1.8e-78;
Matches 375; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
QY 400 CGAGGCTCCCTGCGCGCGGCTGCTGTCATGGACTCGTGAGAGCTCGCTTCCCGCG 459
DB 389 CGAGGCTCCCTGCGCGCGGCTGCTGTCATGGACTCGTGAGAGCTCGCTTCCCGCG 330
QY 460 CGGACCCCTTCCTGCAGGGGTCCACGTCGAGGACCGCGGCTCGGACACCCACCCCGCG 519
DB 329 CGGACCCCTTCCTGCAGGGGTCCACGTCGAGGACCGCGGCTCGGACA-CCACCCCGCG 271
QY 520 CGGGGACCTGCGCTGGGTGCGCCCTTAACCCGGGCGGTAGCTGCTTAAGATGCGGAAGTG 579
DB 270 CGGGGACCTGCGCTGGGTGCGCCCTTAACCCGGGCGGTAGCTGCTTAAGATGCGGAAGTG 211
QY 580 TCGGTGCGGACACGCGAACCACAAATCCGCGCTGCCGACCTCTGACCCCGCGCC 639
DB 210 TCGGTGCGGACACGCGAACCACAAATCCGCGCTGCCGACCTCTGACCCCGCGCC 151
QY 640 CACGGGACGACAGCTGGGCTCCCGACGCGAGCGGCTCGCGGACACCGGTGGGTGC 699
DB 150 CACGGGACGACAGCTGGGCTCCCGACGCGAGCGGCTCGCGGACACCGGTGGGTGC 91
QY 700 GAAACGAGGACCTTTGTAAACGCGCTGTTGCTCTTTTGAAGAAACAAAGTAAATG 759
DB 90 GAAACGAGGACCTTTGTAAACGCGCTGTTGCTCTTTTGAAGAAACAAAGTAAATG 31
QY 760 TGTAAACTGCTGAAAA 777
DB 30 TGTAAACTGTCAAAAA 13

RESULT 14
BF570483/c
LOCUS
DEFINITION
602186293T1 NIH_MGC_45 Homo sapiens cDNA clone IMAGE:4310518 3',
mRNA sequence.
ACCESSION
BF570483
VERSION
BF570483.1 GI:11644195
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 457)
NIH-MGC <http://mgi.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Linehan
cDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLCM185 row: j column: 23
High quality sequence start: 30
High quality sequence stop: 448.
Location/Qualifiers

FEATURES
source

1..457
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4310518"
/tissue_type="renal carcinoma (ascites)"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_45"
/note="Organ: Kidney; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

ORIGIN

Query Match 31.0%; Score 303.6; DB 2; Length 457;
Best Local Similarity 91.7%; Pred. No. 6.9e-64;
Matches 378; Conservative 0; Mismatches 24; Indels 10; Gaps 5;
QY 283 GCGAAGACGCGCTGGGGGACCCCTTGGTCCGCTCTGCGGCTCGCTTAGAGTCCCCGAAA 342
DB 447 GAGAACGCGCATGGGGGAAACCATTTGGTCCGCTCTGCGGCTCGCTTAGAGTCCCCGAAA 388
QY 343 AGGAGCAGCGGCGCGAAGCGGCCAGGCTGGGCCAGGATCTAGAAAGACTGCTT-GGCGC 401
DB 387 AGGAGCAGCGGCGCGAAGCGGCCAGGCTGGGCCAGGATCTAGAAAGACTGCTTAGAGCGC 328
QY 402 AGGCTCCCTTCCCGCGCGGCGC---TGCTGTCTAGGACT--CGTGGAGAGCTCGCTTCC 455
DB 327 AGGCTCCCTTCCCGCGCGGCGCTAGCTCGATCATGGACTAGCTCGAGAGCTCGCTTCC 268
QY 456 CGCGCGGACCTTCC--TGCAGGGTCCACGTCAGGACACCGCGGCTCGGACACCCAC 513
DB 267 CGCGCGGACCTTCCATGTCAGGAGGTCCACGTCAGGACACCGCGGCTCGGACACCCAC 208
QY 514 CCCCGCGCGGACCTTCCCTGCGTGGTCCCC-TTAAACCGCGGCTAGCTGCTTAAAGTGS 572
DB 207 CCCCGCGCGGACCTTCCCTGCGTGGTCCCCATTAAACCGCGGCTAGCTGCTTAAAGTGS 148
QY 573 CGAAGTCTCGGTCGCGAACAACCGGAAACCCCAAAATCCCGCTGCCGACCTCTGACCC 632
DB 147 CGAAGTCTCGGTCGCGAACAACCGGAAACCCCAAAATCCCGCTGCCGACCTCTGACCC 88
QY 633 CGCGGCGGACGAGGACGACAGCTGGGCTTCCGAGCGCGGCGGCTGCGCGG 684
DB 87 CGCGGCGGACGAGGACGAGCTGGGCTTCCCGAGCGCGGCGGCTGCGCGG 36

RESULT 15

AL198389/c
LOCUS
DEFINITION
q162f10.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:1861099 3',
mRNA sequence.
ACCESSION
AL198389
VERSION
AL198389.1 GI:3750995
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 263)
NCI/NIHNS-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project

(CGAP/BTGP), Tumor Gene Index
Unpublished (1998)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 443 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 238.

FEATURES
source
1..263
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1861099"
/tissue_type="anaplastic oligodendroglioma"
/lab_host="DH10B"
/clone_lib="NCI CGAP Brn25"
/note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCATCTGAGTGGGAGCGCGCATAGGTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
Library is normalized, and was constructed by Bento
Soares and M. Fatima Bonaldo."

ORIGIN
Query Match 26.8%; Score 263; DB 1; Length 263;
Best Local Similarity 100.0%; Pred. No. 6.3e-54;
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 521 CGGGACCTGCCCTGGTGGTCCCTTAACCCGGCGGTAGCTCGTTAAGATGGGAAGTGT 580
DB 263 CGGGACCTGCCCTGGTGGTCCCTTAACCCGGCGGTAGCTCGTTAAGATGGGAAGTGT 204
QY 581 CGGGTCCGGAACACGCGAAACCCCAATCCCGCTGCCCGACCTCTGTACCCCGGCCCC 640
DB 203 CGGGTCCGGAACACGCGAAACCCCAATCCCGCTGCCCGACCTCTGTACCCCGGCCCC 144
QY 641 ACGGACGACAGACTGGGCTCCCGACGCGCGCTGCGGGACACCGGTGCGTGGCG 700
DB 143 ACGGACGACAGACTGGGCTCCCGACGCGCGCTGCGGGACACCGGTGCGTGGCG 84
QY 701 AACGAGGACCTTTGTACGCGACGTGTGCTCTTTTGAATAACAGATTAATGT 760
DB 83 AACGAGGACCTTTGTACGCGACGTGTGCTCTTTTGAATAACAGATTAATGT 24
QY 761 GTTAAACTGCTGAAAAGCTTGC 783
DB 23 GTTAAACTGCTGAAAAGCTTGC 1

Search completed: October 9, 2005, 03:24:42
Job time : 3173 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 8, 2005, 23:39:32 ; Search time 571 Seconds
(without alignments)

10159.976 Million cell updates/sec

Title: US-10-829-118-1

Perfect score: 980

Sequence: 1 ctccgcggcgaacacgcctg.....ctgcaccacggaatgctga 980

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq_16Dec04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002as:*
7: Geneseqn2002bs:*
8: Geneseqn2003as:*
9: Geneseqn2003bs:*
10: Geneseqn2003cs:*
11: Geneseqn2003ds:*
12: Geneseqn2004as:*
13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	980	100.0	980	6	AAD24180 Human sit
2	980	100.0	2469	6	AAD24181 Human sit
C 3	124.4	12.7	4338	3	AAAS7197 Human sub
C 4	124.4	12.7	4338	6	AAD24182 Human sit
C 5	124.4	12.7	4338	10	ADG32883 Human DNA
C 6	124.4	12.7	4338	13	ADR25092 Breast ca
C 7	64	6.5	114955	2	AAAS3491 Human ade
C 8	60.6	6.2	114955	2	AAAS3491 Human ade
C 9	53	5.4	110000	4	AAI99683_16 Continuation (17 o
C 10	52.6	5.4	4667	12	ADQ22939 Human sof
C 11	51.2	5.2	2000	8	ADA71938 Rice gene
C 12	50.8	5.2	2307	6	ABK32842 DNA encod
C 13	50.8	5.2	4542	10	ADC30607 Human nov
C 14	50.8	5.2	4650	10	ADC32422 Human nov
C 15	50.2	5.1	24000	3	AAAS8551 Human dua
C 16	49.8	5.1	849	6	ABQ29961 Oligonuc
C 17	49.8	5.1	849	6	ABQ29960 Oligonuc
C 18	48.8	5.0	771	6	AAD41042 Human COL
C 19	48.8	5.0	771	6	ABT03863 Human Goo
C 20	48.8	5.0	771	6	ABT03864 Human Goo

C 21	48.8	5.0	771	9	ACD13467 Human Pol
C 22	48.8	5.0	771	9	ACD13466 Human GPB
C 23	48.8	5.0	771	9	ADA97869 Human tum
C 24	48.8	5.0	771	9	ADA97868 Human tum
C 25	48.8	5.0	955	6	AAD41041 Human 955
C 26	48.8	5.0	955	6	ABT03862 Human Goo
C 27	48.8	5.0	955	9	ACD13465 Human pol
C 28	48.8	5.0	955	9	ADA97867 Human tum
C 29	48.8	5.0	40668	6	ABQ88150 Human ost
C 30	48.4	4.9	4225	10	ABV75290 Mouse SCA
C 31	48.2	4.9	1187	12	ACH91930 Human gen
C 32	48	4.9	38847	11	ACN44426 Human gen
C 33	47.6	4.9	989	3	AAA02539 Human col
C 34	47.4	4.8	154746	6	AAD25519 Human her
C 35	47.4	4.8	154746	6	AAD25519 Human her
C 36	46.8	4.8	1535	9	AAD58542 Human thy
C 37	46.8	4.8	1609	13	ADR07763 Full leng
C 38	46.8	4.8	2571	10	ADC63369 Human CDN
C 39	46.8	4.8	2571	10	ABX93989 cDNA enco
C 40	46.8	4.8	2571	12	ADN49355 Human kin
C 41	46.8	4.8	2856	9	ADA38100 Thymidine
C 42	46.8	4.8	2856	10	ADC39135 Novel hum
C 43	46.8	4.8	2856	11	ADM10601 Human thy
C 44	46.8	4.8	2856	12	ADH42558 Novel hum
C 45	46.8	4.8	2893	13	ACN37638 Tumour-as

ALIGNMENTS

RESULT 1
AAD24180
ID AAD24180 standard; DNA; 980 BP.
XX
AC AAD24180;
XX

DT 07-MAY-2002 (first entry)

XX Human site-1 protease promoter partial sequence.

XX Site-1 protease, S1P; promoter; transcription; obesity; type II diabetes;
KW sterol regulatory element-binding protein; SREBP; hypercholesterolaemia;
KW dyslipidaemia; atherosclerosis; cardiovascular disease; human; ds.
XX

OS Homo sapiens.

XX
PN WO200200873-A1.
XX

PD 03-JAN-2002.

PF 19-JUN-2001; 2001WO-SE001386.

XX 27-JUN-2000; 2000SE-00002417.

PR (BIOV-) BIOVITRUM AB.

PI Abrahamsen L, Ekblom J, Forsgren M, Hoerling J, Johansson P;

CC is a fragment of human S1P promoter

XX	SQ	Sequence	980 BP; 175 A; 313 C; 326 G; 166 T; 0 U; 0 Other;
		Query Match 100.0%; Score 980; DB 6; Length 980; Best Local Similarity 100.0%; Pred. No. 6.4e-235; Matches 980; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Qy	1	CTCCGCGCGGAACACGCTCGGCGACTCCATTTCGGGGCTGTTTACTCCCAACTCTCGCGAG	60
Db	1	CTCCGCGCGGAACACGCTCGGCGACTCCATTTCGGGGCTGTTTACTCCCAACTCTCGCGAG	60
Qy	61	ACTGGGGCGCGCGGCACGACGAGGCGCCACAGCTGGGAGCCTCAGCTCGCGCGACCCAGCGT	120
Db	61	ACTGGGGCGCGCGGCACGACGAGGCGCCACAGCTGGGAGCCTCAGCTCGCGCGACCCAGCGT	120
Qy	121	GCCCTGTCTGTCCTCGCGCTCCCGGGCTTGGGTGGCGCTCTGGACGCGCTGGGGAGCGG	180
Db	121	GCCCTGTCTGTCCTCGCGCTCCCGGGCTTGGGTGGCGCTCTGGACGCGCTGGGGAGCGG	180
Qy	181	GACCACCGCGGAGGATGACGAAAGTGTCTCGCACATTTTCGGCGCGGGGGCGCGGTGG	240
Db	181	GACCACCGCGGAGGATGACGAAAGTGTCTCGCACATTTTCGGCGCGGGGGCGCGGTGG	240
Qy	241	CAGGTTGAACGGAGGGCGCTGGCCAGCAGCTGCCAGGCGGCGAGAACCGCTGGGGG	300
Db	241	CAGGTTGAACGGAGGGCGCTGGCCAGCAGCTGCCAGGCGGCGAGAACCGCTGGGGG	300
Qy	301	AACCTTTGGTCCGCTCTGCGCGTCTAGGATCCCGGAAAAGGAGCACGGGCGCGAAA	360
Db	301	AACCTTTGGTCCGCTCTGCGCGTCTAGGATCCCGGAAAAGGAGCACGGGCGCGAAA	360
Qy	361	CGGCGCAGGCTGGCCAGGATCTAGAAAGCTGCTGGCGCAGACTCCCTGCCCGCCCGGG	420
Db	361	CGGCGCAGGCTGGCCAGGATCTAGAAAGCTGCTGGCGCAGACTCCCTGCCCGCCCGGG	420
Qy	421	GCCTGCTGTCAATGGACTCGTGGAGAGCTCGCTTCCCGCGCGGACCCCTTCCTGCGAGGGTC	480
Db	421	GCCTGCTGTCAATGGACTCGTGGAGAGCTCGCTTCCCGCGCGGACCCCTTCCTGCGAGGGTC	480
Qy	481	CACGTCCAGGCAACCGGGGCTCGGACACCCCAACCCCGCGCGCGGACCTGCGCTGGGTGC	540
Db	481	CACGTCCAGGCAACCGGGGCTCGGACACCCCAACCCCGCGCGCGGACCTGCGCTGGGTGC	540
Qy	541	CCCTTAACCCGGGCGGTAGCTCTTGAATGGCGAAGTGTCCGGTCCGGACACGCGAAA	600
Db	541	CCCTTAACCCGGGCGGTAGCTCTTGAATGGCGAAGTGTCCGGTCCGGACACGCGAAA	600
Qy	601	CCCCAAATCCCGCTGCGCCGACTCTCAGCCCGGCCCCACGGGACGACAGACTGGGCC	660
Db	601	CCCCAAATCCCGCTGCGCCGACTCTCAGCCCGGCCCCACGGGACGACAGACTGGGCC	660
Qy	661	TCCCGACGCGAGCGCTCGCGGACACCGGTGCGTGCAGAAAACGGAGGACCTTTGTAAAC	720
Db	661	TCCCGACGCGAGCGCTCGCGGACACCGGTGCGTGCAGAAAACGGAGGACCTTTGTAAAC	720
Qy	721	GCCAGGTTTGTCTCTTTTGAAGAAAACAGAAATAAATGTTGTTAACTGTCTGAAAAGCT	780
Db	721	GCCAGGTTTGTCTCTTTTGAAGAAAACAGAAATAAATGTTGTTAACTGTCTGAAAAGCT	780
Qy	781	TGCGCGCTTAAAGATGTCTGGGTGACTTAGATGCTTAGATCAGTTTGTGTTTCAATGTAAA	840
Db	781	TGCGCGCTTAAAGATGTCTGGGTGACTTAGATGCTTAGATCAGTTTGTGTTTCAATGTAAA	840
Qy	841	TGGACCAAGCCCGGACTCCGTTACGGCACTAGCAGGGGACTGAAAGCGCTTTCAGGTACTGC	900
Db	841	TGGACCAAGCCCGGACTCCGTTACGGCACTAGCAGGGGACTGAAAGCGCTTTCAGGTACTGC	900
Qy	901	TGTTGGGCGGTGATGCGCTACAGCCGATCAGACAGTTTGTGTTTGTCTTCTGGAACCTTGACA	960
Db	901	TGTTGGGCGGTGATGCGCTACAGCCGATCAGACAGTTTGTGTTTGTCTTCTGGAACCTTGACA	960
Qy	961	CTGCACCAACCGTAAATGCTGA	980

Db 961 CTGCACCACGGTAATGCTGA 980

RESULT 2
AAD24181
ID AAD24181 standard; DNA: 2469 BP.

XX
DE Human site-1 protease promoter region.
XX
KW Site-1 protease; SLP; promoter; transcription; obesity; type II diabetes;
KW sterol regulatory element-binding protein; SRBP; hypercholesterolaemia;
KW dyslipidaemia; atherosclerosis; cardiovascular disease; human; ds.
KW

XX
PN
XX
PD

XX
PI
Abrahamsen L. Ekblom J. Forsgren M. Hoerling J. Johansson P;

XX Human site-1 protease promoter region for identifying agents capable of
PT inhibiting the promoter activity useful in treating medical conditions
PT such as obesity, diabetes, atherosclerosis and hypercholesterolemia.

PS Claim 1; Page 15-16; 36pp; English.

CC The invention relates to human site-1 protease (S1P) promoter region. The
CC promoter sequence is useful for identification of compounds that inhibit
CC transcription of S1P, which in turn results in inhibition of sterol
CC regulatory element-binding protein (SREBP) pathway. The compound
CC identified is useful for the treatment of medical conditions related to
CC obesity, type II diabetes, hypercholesterolaemia, dyslipidaemia,
CC atherosclerosis and other cardiovascular diseases. The present sequence
CC is human S1P promoter region

Sequence 2469 BP; 541 A; 704 C; 719 G; 505 T; 0 U; 0 Other;

Query Match	100.0%;	Score 980;	DB 6;	Length 2469;
Best Local Similarity	100.0%;	Pred. No. 8.3e-235;		

Qy 1 CTCCGGCGGAAACACGCCCTGGGGCACTCCATTGGGGCTGTTTACTCCCCAACTCTCGCGAG 60

Dh 153 CTCCGGCGGCGAAACACGCCCTGGGGCACTCCATTGGGGCTGTTTACTCCCCAACTCTCGCGAG 412

Qy 961 CTGCACCACGGTAATGCTGA 980

593 CAGGGTGGAAAGCGAGGGGGCTGGCCAGCGAGCTGCCAGCGCGGAGAAACGGCGCTGGGG 652
301 AACCTTGGTCCGCTCTGGCGCTCGCTCTAGGATCCCGAAAAAGAGACAGCGCGCGAAA 360
653 AACCTTGGTCCGCTCTGGCGCTCGCTCTAGGATCCCGAAAAAGAGAGACAGCGCGCGAAA 712
361 GCGGCCAGGCTGGCCAGAGATCTAGAAAGACTGCGCTGGCGAGCTCCCTGCCCCCGCG 420
713 GCGGCCAGGCTGGCCAGAGATCTAGAAAGACTGCGCTGGCGAGCTCCCTGCCCCCGCG 772
421 GCTGCTGTCATGAGCTCGTGGAGAGCTCGCTCCCGCGGAGCCCTCTCGAGGGTTC 480
773 GCTGCTGTCATGAGCTCGTGGAGAGCTCGCTCCCGCGGAGCCCTCTCGAGGGTTC 832
481 CAGCTCAGGACCGCGGCTCGGACACCCACCCCGCGCGGACCTGCCCTGGGTGC 540
833 CAGTCCAGGACCGCGGCTCGGACACCCACCCCGCGCGGACCTGCCCTGGGTGC 892
541 CCTTAAACCGCGGCTAGTCTGTTAAGATGGCGAAGTGTCCGGTCCGGAACACGCGAAA 600
893 CCTTAAACCGCGGCTAGTCTGTTAAGATGGCGAAGTGTCCGGTCCGGAACACGCGAAA 952
601 CCCCAATCCGCTCCGCGACTCTGACCCCGCGCGCCACCGGACGACAGCTGGGC 660
953 CCCCAATCCGCTCCGCGACTCTGACCCCGCGCGCCACCGGACGACAGCTGGGC 1012
661 TCCGACGCGCAGCGCTGCCGCGGACACCGGTGCGTGGAAACGAGGACCTTTGTAAC 720
1013 TCCGACGCGCAGCGCTGCCGCGGACACCGGTGCGTGGAAACGAGGACCTTTGTAAC 1072
721 GGCACGTGTTGCTTTTTGAAAAAACAAGATAAATGTGTTAAACTGTCTGAAAGCT 780
1073 GGCACGTGTTGCTTTTTGAAAAAACAAGATAAATGTGTTAAACTGTCTGAAAGCT 1132
781 TGGCGCTAAAGATGTCTGGGTGACTTAGTACTAGGATCAGTTGTTTCAATGATAA 840
1133 TGGCGCTAAAGATGTCTGGGTGACTTAGTACTAGGATCAGTTGTTTCAATGATAA 1192
841 TGGACGACCGGACTCCGTACGGCAGTACGAGGAGCTGAAAGCGCTTTCAGGTACTGC 900
1193 TGGACGACCGGACTCCGTACGGCAGTACGAGGAGCTGAAAGCGCTTTCAGGTACTGC 1252
901 TGGTGGCGGTGATGCGCTACAGGCCGATCAGACAGTTTGTGTTCTTCTGGAACCTTGACA 960
1253 TGGTGGCGGTGATGCGCTACAGGCCGATCAGACAGTTTGTGTTCTTCTGGAACCTTGACA 1312
961 CTGACACGAGTAATGCTGA 980
1313 CTGACACGAGTAATGCTGA 1332

RESULT 3

AAH57197/c
ID AAA57197 standard; cDNA; 4338 BP.

XX AC
XX AAH57197;

XX DT
XX 03-OCT-2000 (first entry)

XX DE
XX Human subtilisin-kexin isoenzyme 1 cDNA.

XX KW
XX Human; subtilisin-kexin isoenzyme 1; SKI-1; antilipaeamic; cytostatic;
KW vasotropic; SKI-1 inhibitor; hypercholesterolaemia; liver steatosis;
KW Ras-dependent cancer; restenosis; amyloid protein formation;
KW pro-brain-derived neurotrophic factor; proBDNF;
KW sterol-regulatory element-binding protein; SREBP; ds.

XX OS
XX Homo sapiens.

XX FH
XX Key Location/Qualifiers
FT CDS 497..3655
FT /*tag= a
FT /product= "SKI-1"

XX WO200026348-A2.
XX 11-MAY-2000.
XX 04-NOV-1999; 99WO-CA001058.
XX 04-NOV-1998; 98CA-02249648.
XX (RECL-) INST RECH CLINIQUES MONTREAL.
XX Seidah N, Chretien M, Marcinkiewicz M, Laaksonen R, Davignon J;
XX WPI; 2000-365601/31.
XX P-PSDB; AAB06334.
XX Novel soluble proteic fragment of subtilisin-kexin isoenzyme for
XX producing a polypeptide useful for treating hypercholesterolemia, liver
XX steatosis and amyloidosis, comprises a specific amino acid sequence.
XX Example 1; Page 56-60; 119pp; English.
XX The present sequence encodes human subtilisin-kexin isoenzyme 1 (SKI-1),
XX a type-1 membrane-bound proteinase. The nucleotide sequence was isolated
XX from the human neuronal cell line IMR-32 by RT-PCR using active site
XX degenerate primers. SKI-1 cleaves at a specific threonine residue within
XX the N-terminal segment of human pro-brain-derived neurotrophic factor
XX (proBDNF). It is also capable of cleaving sterol-regulatory element-
XX binding proteins (SREBPs), which function to control lipid biosynthesis
XX and uptake in animal cells. Peptides which bind to and are cleaved by SKI
XX -1 may be used for monitoring SKI-1 activity, for screening inhibitors of
XX SKI-1 activity, or for screening enhancers of SKI-1 activity. Proteic
XX fragments of SKI-1 which bind to the SKI-1 catalytic site may be used as
XX inhibitors of SKI-1 activity. They may be used to treat diseases
XX involving overexpression of SKI-1 or SKI-1 substrate. Such diseases
XX include hypercholesterolaemia, high levels of fatty acids, lipids or
XX farnesyl pyrophosphate, liver steatosis, Ras-dependent cancer, restenosis
XX and amyloid protein formation
XX Sequence 4338 BP; 1085 A; 1051 C; 1160 G; 1042 T; 0 U; 0 Other;
XX
XX Query Match 12.7%; Score 124.4; DB 3; Length 4338;
XX Best Local Similarity 99.2%; Pred. No. 8.5e-21;
XX Matches 125; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CTCGCGCGGGAACACGCGCTGGGCACTCCATTTCGGGGCTGTTTACTCCCAACTCTCGCGAG 60
Db 126 CTCGCGCGGGAACACGCGCTGGGCACTCCATTTCGGGGCTGTTTACTCCCAACTCTCGCGAG 67
QY 61 ACTGGCGCGCGCGGCGAGGCGCCACAGCTGGGAGCTCAGCTCCGCGGACCCAGCGCT 120
Db 66 ACTGGCGCGCGCGGCGAGGCGCCACAGCTGGGAGCTCAGCTCCGCGGACCCAGCGCT 7
QY 121 GCCCTG 126
Db 6 GCCCTG 1
RESULT 4
AAD24182/c
ID AAD24182 standard; DNA; 4338 BP.
XX AC
XX AAD24182;
XX 07-MAY-2002 (first entry)
XX Human site-1 protease DNA.
XX Site-1 protease; S1P; promoter; transcription; obesity; type II diabetes;
KW sterol regulatory element-binding protein; SREBP; hypercholesterolaemia;
KW dyslipidaemia; atherosclerosis; cardiovascular disease; human; ds.
XX OS
XX Homo sapiens.

QY 476 GGGTCCACGTCACGGCAGCGGCTCGGACACCCACCGCGCGGACCTGCGCCCTG 535
Db 105261 CGCGCNHNHNSCGGCGCGCGCGCGCGCGCNHNHNSCGGCGCGCGCGCGCGCG 105202
QY 536 GGTGCCCCCTAAACCGCGGCGGTAGCTCGTTAAGATGGCGAAGTGTCCGGTCCGGAACAG 595
Db 105201 CCNHNHNSCGGCGCGCGCGCGCGCGCCVNNHNHNSCGGCGCGCGCGCGCGCG 105142
QY 596 CGAAACCCCAATTCGCGCTGCGCGACCTCTGACCCCCCGCGCCCAACGGGACAGACT 655
Db 105141 CVGNHNHNSCGGCGCGCGCGCGCGCGCCVGNHNHNSCGGCGCGCGCGCGCGCG 105082
QY 656 GGGCTCCCGACGCGCAGCGCG 677
Db 105081 GCCVGGCNHNHNSCGGCGCG 105060

RESULT 8

AX53491
ID AX53491 standard; DNA; 114955 BP.

AX
AC AAX53491;

XX
XX 05-JUL-1999 (first entry)

XX Human adenosine A1 receptor antisense oligonucleotide fragment.

XX Antisense oligonucleotide; multiple target; antisense treatment;
KW impaired respiration; inflammation; lung disease;
KW pulmonary vasoconstriction; inflammation; allergic rhinitis;
KW acute asthma; allergy; asthma; impeded respiration;
KW respiratory distress syndrome; pain; cystic fibrosis;
KW pulmonary hypertension; pulmonary vasoconstriction; emphysema;
KW chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;
KW colon cancer; breast cancer; lung cancer; pancreatic cancer;
KW hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;
KW prostate cancer; ss.

XX Synthetic.

XX WO9913886-A1.

XX 25-MAR-1999.

XX 17-SEP-1998; 98WO-US019419.

XX 17-SEP-1997; 97US-0059160P.

XX 09-JUN-1998; 98US-00093972.

XX (UYEC-) UNIV EAST CAROLINA.

XX Nyce JW;

XX WPI; 1999-229400/19.

XX New antisense oligonucleotides used in treatment of, e.g. pulmonary
PT vasoconstriction.
XX Disclosure; Page 37; 120pp; English.

XX The specification describes antisense oligonucleotides (AAX52869-X55271)
CC directed against at least 2 mRNAs selected from target genes, coding and
CC non-coding regions of RNAs corresponding to target genes, gene initiation
CC codons, genomic flanking regions, intron-exon borders, the 5'-end, the 3'
CC end and the juxta-section between coding and non-coding regions and all
CC segments of RNAs encoding proteins associated with one or more diseases,
CC conditions or mixtures. The antisense oligonucleotides may be derived
CC from sequences AAX55272-74. These multiple target oligonucleotides
CC (specifically AAX55180-271) can be used for the antisense treatment of
CC diseases and conditions. Typical diseases and conditions are those
CC associated with impaired respiration and inflammation, including lung
CC diseases, pulmonary vasoconstriction, inflammation, allergic rhinitis,

CC acute asthma, allergies, asthma, impeded respiration, respiratory
CC distress syndrome, pain, cystic fibrosis, pulmonary hypertension,
CC pulmonary vasoconstriction, emphysema, chronic obstructive pulmonary
CC disease (COPD), and cancers such as leukemias, lymphomas, carcinomas e.g.
CC colon cancer, breast cancer, lung cancer, pancreatic cancer,
CC hepatocellular carcinoma, kidney cancer, melanoma, hepatic metastases, as
CC well as all types of cancers which may metastasize or have metastasized
CC to the lungs, including breast and prostate cancer

XX Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T; 0 U; 21427 Other;

QY Query Match 6.2%; Score 60.6; DB 2; Length 114955;
Db Best Local Similarity 32.1%; Pred. NO. 0.0002;
Matches 210; Conservative 62; Mismatches 377; Indels 6; Gaps 1;

QY 32 CGGCGCTGTTTACTCCCAACTCTCGGAGACTGGGGCGCGCGCCAGGCACACG 91
Db 104692 CBGGCCBGCGCGCGCNNDNNCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 104751
QY 92 TGGGAGCCTCAGCTCCGCGCAGCCAGCTGCTGTCTGCTCCGCGCTCCCGGGCTTGC 151
Db 104752 GCSNNDNNCCG 104811
QY 152 GTGCGCGCTCTGGACCGCTGGCGAGCGCGACACCGCGGGAGGATGACGAAGTGTCTC 211
Db 104812 CBGGGSGNNDNNCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 104871
QY 212 GCGACATTTGCG 271
Db 104872 NNDNNCCG 104931
QY 272 GCTGCCAGCGCGCGAGAACGCGCTGCGGGGAAACCCCTTGTGCTCTGCGCGCTCTAG 331
Db 104932 GCGGCG 104991
QY 332 GATCCCCGAAAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 391
Db 104992 GGGCGSGNNDNNBGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 105051
QY 392 TGCCTGGCGAGGCTCCCTGCCCCCGCGG-----GCTGTGTCTATGACTCTGTGAGA 445
Db 105052 CCG 105111
QY 446 GCTGCTTCCCG 505
Db 105112 GCG 105171
QY 506 CACCCACACCG 565
Db 105172 GCG 105231
QY 566 AAGATGGGGAAGTGTCCGTCGGAACACGCGAAACCCCAATTCGCGCTGCGCGACCTC 625
Db 105232 GCG 105291
QY 626 CTGACCCCG 680
Db 105292 CGGCG 105346

RESULT 9

AAI99683_16/c

Continuation (17 of 44) of AAI99683 from base 1600001 (Mycobacterium tuberculosis strain
WP Sequence split into 44 fragments LOCUS AAI99683 Accession AAI99683

WP	Fragment Name	Begin	End
WP	AAI99683_00	1	110000
WP	AAI99683_01	100001	210000
WP	AAI99683_02	200001	310000
WP	AAI99683_03	300001	410000
WP	AAI99683_04	400001	510000
WP	AAI99683_05	500001	610000
WP	AAI99683_06	600001	710000
WP	AAI99683_07	700001	810000

XX 29-JUN-2000; 2000US-0215164P.
 PR 10-AUG-2000; 2000US-0224457P.
 XX (ANAD-) ANADYS PHARM INC.
 PA Moore J, Buurman ET, Desilva T, Harris S, Komarnitsky S;
 PI Mendillo M, Moore D, McCoy M, Sanderson K, Haq T, Zhu S, Long P;
 PI Davidov E, Thompson CW;
 XX WPI; 2002-147962/19.
 DR P-PSDB; AAU82954.
 XX Screening candidate antifungal compound for interaction with essential
 PT protein, modulation of essential protein activity, binding to essential
 PT protein, by contacting protein with test compound and determining
 PT effects.
 XX Disclosure; Fig 80; 522pp; English.
 XX The invention describes a method of screening a candidate antifungal
 CC compound for interaction with essential proteins (EP) or for modulation
 CC of EP activity e.g. fungal gene transcription. The proteins tested in the
 CC invention include RPC34, POP3, TPA2, NAB2, MPT1, MTR2, BOS1, POL30, RSA2,
 CC SQT1, MTW1, TFB1, SPC38, BFR2, RNAL, GCD7, SKI16, NIP1, LCP5, NCE103,
 CC ECO1, ORC2, CNS1, YPD1, TIM10 and SRB4 from *S. cerevisiae*, *C. albicans*
 CC and human homologues. The method involves contacting a culture with one
 CC or more test compounds and determining the effects on the growth or
 CC viability of the culture of cells which preferably comprises fungal cells
 CC or yeast cells. Preferably the identified compounds interact with, or
 CC modulate (preferably inhibit) the activity of *C. albicans* EP. The inhibitor
 CC compounds identified by the method are useful for preventing or
 CC inhibiting fungal, particularly *C. albicans* growth in culture or in a
 CC mammal. The antifungal agents interact with essential fungal elements
 CC that can be used to treat fungal infection by preventing the growth and
 CC preferentially killing the fungi, but does not inhibit the biological
 CC activity of mammalian homologues. This sequence encodes a target protein
 CC used to test the antifungal compounds, described in the method of the
 CC invention
 XX
 SQ Sequence 2307 BP; 359 A; 969 C; 724 G; 255 T; 0 U; 0 Other;
 Query Match 5.2%; Score 50.8; DB 6; Length 2307;
 Best Local Similarity 47.9%; Pred. No. 0.018;
 Matches 181; Conservative 0; Mismatches 192; Indels 5; Gaps 1;
 QY 134 CGCGCTCCGCGGCTTGGTGGCGGCTTGGACGCGGTGGGACGCGGACCGCGGGA 193
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 579 CGCGGCGCGGCGCGCGGCTTGGACGCGGCGCGGCGCGGCGCGGCGCGGCGG 520
 QY 194 GGATGACGACGAGTGTCTCGGACATTTGGCGCGCGGCGCGGCGCGGCGGCGG 253
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 519 GCGGCGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 460
 QY 254 GAGGCGGCGTGGCGGCGGAGTGGCA-----GGCGGCGGAGAACCGCTGGGGGAACCCCTTG 308
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 459 GCGGCGGCTGGGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 400
 QY 309 GTCCGCTCTGCGGCTGCTTAGATCCCGGAAAGAGACGCGGCGGCGGCGGCGGCGG 368
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 399 GCTGCGCTCTGCGGCGGCGGCGGCTAGCTTCGCGGCGGCGGCGGCGGCGGCGGCGG 340
 QY 369 GCTGCGGCGGAGTCTAGAAAGACTGCTGCGGCGGAGCTCCCTCGCGGCGGCGGCGGCTGCTG 428
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 339 GGGGCGGCGGCTGAGGGGGGCGCGGCGGCGCTCGCGGCGGCGGCGGCGGCGGCGGCT 280
 QY 429 TCATGACTCTGTGAGAGAGTCTGCTTCCCGCGGAGCCCTTCCTGCGAGGGGTCCACCTCCA 488
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 279 ACTTGGCGGCGGCGGCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 220
 QY 489 GGACCGCGGCGGCTCGGAC 506
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 219 TGGCGCGGCGGCTCGGCG 202

RESULT 13
 ADC30607/C
 ID ADC30607 standard; cDNA; 4542 BP.
 XX
 AC ADC30607;
 XX
 DT 18-DEC-2003 (first entry)
 XX
 DE Human novel cDNA sequence, SEQ ID NO: 689.
 XX
 KW Human; diagnostic; drug screening; forensics; gene mapping;
 KW biodiversity assessment; Parkinson's disease; Alzheimer's disease;
 KW neurodegenerative diseases; anaemia; platelet disorder; wound; burns;
 KW ulcers; osteoporosis; autoimmune disease; cancer;
 KW molecular weight marker; food supplement; antiparkinsonian; nootropic;
 KW neuroprotective; antianaemic; anticoagulant; thrombolytic; vulnery;
 KW antiulcer; osteopathic; immunosuppressive; antiinflammatory; cytostatic;
 KW gene therapy; chromosome 20; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO2003029271-A2.
 XX
 PD 10-APR-2003.
 XX
 PF 24-SEP-2002; 2002WO-US030474.
 XX
 PR 24-SEP-2001; 2001US-0324631P.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;
 PI Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;
 PI Haley-Vicente D, Drmanac RT;
 XX WPI; 2003-371981/35.
 DR P-PSDB; ADC31578.
 XX
 PT New polynucleotide and polypeptide useful for diagnosing, preventing or
 PT treating conditions such as neurodegenerative diseases, anemias, platelet
 PT disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
 PT cancer.
 PS Claim 1; SEQ ID NO 689; 1185pp; English.
 XX
 CC The invention relates to 971 novel human cDNA sequences (ADC29919-
 CC ADC30889) and the polypeptides they encode (ADC30890-ADC31860). The
 CC invention also relates to nucleic acid sequences over 99% identical with
 CC the novel human cDNAs. The invention additionally encompasses expression
 CC vectors and host cells comprising a nucleic acid of the invention; the
 CC recombinant production of a polypeptide of the invention; a method of detecting
 CC polynucleotides or polypeptides of the invention; and methods of
 CC identifying a compound which binds to a polypeptide of the invention. The
 CC invention further discloses methods of preventing, treating or
 CC ameliorating a medical condition; kits comprising polynucleotide probes
 CC and/or monoclonal antibodies for carrying out the methods of the
 CC invention; methods for the identification of compounds that modulate the
 CC expression or activity of the polynucleotide and/or polypeptide; and 767
 CC contig sequences corresponding to the cDNA sequences of the invention
 CC (ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628
 CC -ADC33394). The nucleic acids and polypeptides of the invention are
 CC useful in diagnostics, drug screening, forensics, gene mapping, in the
 CC identification of mutations responsible for genetic disorders or other
 CC traits, for assessing biodiversity, and in producing many other types of
 CC data and products dependent on DNA and amino acid sequences. They are
 CC also used for treating diseases such as Parkinson's disease, Alzheimer's
 CC disease and other neurodegenerative diseases, anaemia, platelet
 CC disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
 CC cancer. The nucleic acids may also be used as hybridisation probes or
 CC primers, and in the recombinant production of a protein. The polypeptides

QY 429 TCATGACTCGTGGAGAGTCTCGCTTCCCGCGCGACCCCTTCCTGACAGGGTCCACCTCCA 488
Db 4372 ACCTGGCGGGCGGGCTCCGGCGCGCTCGGGCGCGCCCTCGCGGGCGCGCGCC 4431
QY 489 GGCACCGCGGGCTCGGAC 506
Db 4432 TGGCGCCCGGGCTCCGGC 4449

RESULT 15
ID AAA88551/c
XX AAA88551 standard; DNA; 24000 BP.
AC AAA88551;
DT 22-JAN-2001 (first entry)
XX Human dual-specificity phosphatase-1 (DSP-1) gene.
DE DSP-1; dual-specificity phosphatase-1; human; cell proliferation;
KW cell differentiation; cell survival; cell cycle; dephosphorylation;
KW signal transduction; MAP-kinase; cancer; graft versus host disease;
KW allergy; autoimmune disease; metabolic disease; therapy; chromosome 17;
ds.

OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT exon 42..109
FT /*tag= a
FT /number= 1
FT intron 110..20823
FT /*tag= b
FT /number= 1
FT exon 20824..21034
FT /*tag= d
FT /number= 2a
FT exon 20824..20911
FT /note= "Alternative, extended version of exon 2"
FT /*tag= c
FT intron 20912..22327
FT /*tag= e
FT /number= 2
FT intron 21035..22327
FT /*tag= f
FT /number= 2a
FT exon 22328..23309
FT /*tag= g
FT /number= 3
FT CDS 22420..23016
FT /*tag= h

XX WO200053636-A2.
PN
XX
XX 14-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US006154.
XX
XX 08-MAR-1999; 99US-0123255P.
XX
XX (CEPT-) CEPTYR INC.
XX
XX Luche RM, Wei B;
XX
XX WPI; 2000-579365/54.
DR P-PSDB; AAB19602.
XX
XX New isolated polypeptide having the sequence of dual-specificity
PT phosphatase-1 (DSP-1) is useful for treating a patient with a disorder
PT associated with DSP-1 activity e.g. cancer and autoimmune diseases.
XX
XX Example 1; Fig 3A-J; 74pp; English.

XX The present sequence is that of the human dual-specificity phosphatase-1
CC (DSP-1) gene on chromosome 17. The gene was identified in genomic
CC sequences obtained from an expressed sequence tag database screened with
CC a conserved motif (see AAB19604) of known DSPs. DSP-1 dephosphorylates
CC both phosphothreonine/serine and phosphotyrosine residues in DSP-1
CC substrates such as activated mitogen-activated protein kinase (MAP-
CC kinase). DSP-1 has sequence homology to other MAP-kinase phosphatases. It
CC is expressed at high levels in the human heart, testis and liver, and at
CC lower levels in other tissues. Methods are provided for recombinant
CC production of DSP-1 polypeptides, and for using DSP-1 polypeptides,
CC antibodies and polynucleotides to detect DSP-1 expression, to screen for
CC agents that modulate DSP-1 activity e.g. within a combinatorial library,
CC and for using such agents to modulate cell proliferation, cell
CC differentiation or cell survival, through modulation of pattern of gene
CC expression, apoptosis or cell cycle. In particular, the cell displays
CC contact inhibition of cell growth, anchorage-dependent growth or an
CC altered intercellular adhesion property, or is a cell present in a
CC patient afflicted with a disorder associated with DSP-1 activity, such as
CC cancer, graft-versus host disease, autoimmune disease, allergy, metabolic
CC disease, abnormal cell growth, abnormal cell proliferation and abnormal
CC cell cycle
XX
SQ Sequence 24000 BP; 6038 A; 5505 C; 5602 G; 6855 T; 0 U; 0 Other;

Query Match 5.1%; Score 50.2; DB 3; Length 24000;
Best Local Similarity 48.2%; Pred. No. 0.05;
Matches 171; Conservative 0; Mismatches 183; Indels 1; Gaps 1;
QY 186 CCGCGGGAGGATGGACGAAGGTGCTGCGACATTTTCGCGCGGGCGGGCGGGCAGGG 245
Db 435 CGCGCGCGAGGCTGCTGGAGCCCGGGGCACAGGGACAGGGTCCGCGGGCGCGCGCG 376
QY 246 TGGAGCGGAGGGCGTGGCCAGCGAGCTCCAGCGCGCGAGAACCGCTGGGGGAACCC 305
Db 375 CGCGGGGGAGCGCGAGGGGGCGCACGCTCAGGGCCCGCGAGGGTCCGCTGCCGAAGT 316
QY 306 TTGGTCCGCTCTGCGCGTCTAGGATCCCCGAAAGAGACACGGGGCGGAAAGCGGC 365
Db 315 TCGGGCTCTCTCTC-CGCGGGCGCCCGAAGAGTCAAGGACCGCCAGGGAGCAGTGGG 257
QY 366 CAGGCTGGGCCAGGATCTAGAAAGACTGCTCGCGCAGCGCTCCCTGCCCGCGGGCCTG 425
Db 256 GCGCGCGGCACAAAGTTTCCAGAGGCGGGGTGCGCGGGGCTCTGGGTCCCCCTGGCCC 197
QY 426 CTGTATGACTCGTGGAGAGCTCGCTTCCCGCGGAGACCTTCTCTGCAGGGGTCCACGT 485
Db 196 AGGGACCCGAGCGGAAGTCCCGCCCTCCCGCCCGCCACCCCAACACCTTCCCCACAACCC 137
QY 486 CCAGGCACCGCGGGCTCGGACACCCCGCGCGGGCGACCTGCGCTCCCTGGGTGC 540
Db 136 CCACCCCGGGCTGCGGGTCCCTTACCGGGCGGGCGCGGGTTC 82

Search completed: October 9, 2005, 02:31:41
Job time : 575 secs

This Page Blank (uspto)

This Page Blank (uspto)

This Page Blank (uspto)

Db	241	CAGGTTGGAAGCGGAGGGGCGTGGCCAGCGAGCTGCCAGGCGCGAGAAACGGCGCTGGGG	300
Qy	301	AAACCTTTGGTTCGGCTCTGGCGCTCGCTTAGAGATCCCGAAAGAGACACGGGCGCGAAA	360
Db	301	AAACCTTTGGTTCGGCTCTGGCGCTCGCTTAGAGATCCCGAAAGAGACACGGGCGCGAAA	360
Qy	361	GGGCGCAGGCTGGGCGAGATCTAGAAAGACTTGCCTGGCGGAGGCTCCTTGCCCCCGGG	420
Db	361	GGGCGCAGGCTGGGCGAGATCTAGAAAGACTTGCCTGGCGGAGGCTCCTTGCCCCCGGG	420
Qy	421	GCTGCTGTCAATGAGTGTGGAGAGCTCGCTTCCCGCGGACCCCTTCTCAGGGGTC	480
Db	421	GCTGCTGTCAATGAGTGTGGAGAGCTCGCTTCCCGCGGACCCCTTCTCAGGGGTC	480
Qy	481	CAGGTCCAGGCAACCGGCGCTCGGACACCCCAACCCCGCGGCGACCTGCCCTGGGTGC	540
Db	481	CAGGTCCAGGCAACCGGCGCTCGGACACCCCAACCCCGCGGCGACCTGCCCTGGGTGC	540
Qy	541	CCCTTAAACCCGGCGGTAGCTCGTTAAGATGGCGAAGTGTCCGGTCCGGAAACACCGGAAA	600
Db	541	CCCTTAAACCCGGCGGTAGCTCGTTAAGATGGCGAAGTGTCCGGTCCGGAAACACCGGAAA	600
Qy	601	CCCCAAATCCCGCTGCCGACCTCTTGACCCCGCGGCCCAACCGGACGACACTGGGCC	660
Db	601	CCCCAAATCCCGCTGCCGACCTCTTGACCCCGCGGCCCAACCGGACGACACTGGGCC	660
Qy	661	TCCGACGCGCAGCGCGCTGCCGGGACACCGGTGGGTGGGAAACGGAGACCTTTGTAAAC	720
Db	661	TCCGACGCGCAGCGCGCTGCCGGGACACCGGTGGGTGGGAAACGGAGACCTTTGTAAAC	720
Qy	721	GCCAGCTGTTGCTCTTTTGAAGAAACAAAGAAATAAATGTGTTAACTGTCTGAAAGCT	780
Db	721	GCCAGCTGTTGCTCTTTTGAAGAAACAAAGAAATAAATGTGTTAACTGTCTGAAAGCT	780
Qy	781	TGCGCGCTAAAGATGTCTGGGTGACTTAGATGCTAGGATCAGTTGTTTCAATGTAAA	840
Db	781	TGCGCGCTAAAGATGTCTGGGTGACTTAGATGCTAGGATCAGTTGTTTCAATGTAAA	840
Qy	841	TGACACAGCCCGGACTCCGTACGGCACTAGCAGGGGACTGAAAGCGCTCTTCAGGTACTGC	900
Db	841	TGACACAGCCCGGACTCCGTACGGCACTAGCAGGGGACTGAAAGCGCTCTTCAGGTACTGC	900
Qy	901	TGTTGGGCGGTGATCGCTACAGGCGGATCAGACAGTTTGTCTTCTGGAACCTTGACA	960
Db	901	TGTTGGGCGGTGATCGCTACAGGCGGATCAGACAGTTTGTCTTCTGGAACCTTGACA	960
Qy	961	CTGCACACGGTAATGCTGA	980
Db	961	CTGCACACGGTAATGCTGA	980

RESULT 2
AR527528 LOCUS AR527528 2469 bp DNA linear PAT 08-OCT-2004
DEFINITION Sequence 2 from patent US 6723553.
AUTHORS AR527528
PROMOTER AR527528
PROMOTER AR527528.1 GI:53914633
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 2469)
AUTHORS Abrahamson, L., Ekblom, J., Forsgren, M., Horling, J. and Johansson, P.
TITLE Promoter sequences
JOURNAL Patent: US 6723553-A 2 20-APR-2004;
FEATURES
1..2469 Location/Qualifiers
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Query Match 100.0%; Score 980; DB 6; Length 2469;
Best Local Similarity 100.0%; Pred. No. 2e-196;

Matches	980;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	1	CTCCGCGCGGAAACAGCGCTGGGCACTCCTCATTTCCGGGCTGTTTACTCCCAACTCTCCGAG	60						
Db	353	CTCCGCGCGGAAACAGCGCTGGGCACTCCTCATTTCCGGGCTGTTTACTCCCAACTCTCCGAG	412						
Qy	61	ACTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	120						
Db	413	ACTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	472						
Qy	121	GCCCTGTCTGTCTCCCGCGCTCCCGGGCTTGGGTGCGCGCTCTTGACGCGCGCTGGGCGGCGG	180						
Db	473	GCCCTGTCTGTCTCCCGCGCTCCCGGGCTTGGGTGCGCGCTCTTGACGCGCGCTGGGCGGCGG	532						
Qy	181	GACCAACCGCGGAGGATGGAAGAGGTCTCGCGACATTTGGGCGGCGGCGGCGGCGGCGGCGG	240						
Db	533	GACCAACCGCGGAGGATGGAAGAGGTCTCGCGACATTTGGGCGGCGGCGGCGGCGGCGGCGG	592						
Qy	241	CAGGTTGGAAGCGGAGGGGCGTGGCCAGCGAGCTGCCAGGCGGCGGAGAACCGCTGGGGG	300						
Db	593	CAGGTTGGAAGCGGAGGGGCGTGGCCAGCGAGCTGCCAGGCGGCGGAGAACCGCTGGGGG	652						
Qy	301	AACCTTCTGCTCCGCTCTCGCGCTCTAGGATCCCGGAAAGAGACACGGGCGCGGAAA	360						
Db	653	AACCTTCTGCTCCGCTCTCGCGCTCTAGGATCCCGGAAAGAGACACGGGCGCGGAAA	712						
Qy	361	GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	420						
Db	713	GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	772						
Qy	421	GCTGCTGTCTCATGAGTCTGTTGAGAGCTGCTTCCCGCGGCGGACCTTCTCGACGGGTC	480						
Db	773	GCTGCTGTCTCATGAGTCTGTTGAGAGCTGCTTCCCGCGGCGGACCTTCTCGACGGGTC	832						
Qy	481	CAGGTCCAGGCAACCGGCGCTCGGACACCCCAACCCCGCGGCGGCGGCGGCGGCGGCGG	540						
Db	833	CAGGTCCAGGCAACCGGCGCTCGGACACCCCAACCCCGCGGCGGCGGCGGCGGCGGCGG	892						
Qy	541	CCCTTAAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	600						
Db	893	CCCTTAAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	952						
Qy	601	CCCGAAATCCCGCTGCCGCGCTCTGACCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGG	660						
Db	953	CCCGAAATCCCGCTGCCGCGCTCTGACCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGG	1012						
Qy	661	TCCCGACGCGCAGCGCGCTGCCGCGGACACCGGTGGGTGGGAAACGGAGGACCTTTGTAAAC	720						
Db	1013	TCCCGACGCGCAGCGCGCTGCCGCGGACACCGGTGGGTGGGAAACGGAGGACCTTTGTAAAC	1072						
Qy	721	GCCAGCTGTTGCTCTTTTGAAGAAACAAAGAAATAAATGTGTTAACTGTCTGAAAGCT	780						
Db	1073	GCCAGCTGTTGCTCTTTTGAAGAAACAAAGAAATAAATGTGTTAACTGTCTGAAAGCT	1132						
Qy	781	TGCGCGCTAAAGATGTCTGGGTGACTTAGATGCTAGGATCAGTTGTTTCAATGTAAA	840						
Db	1133	TGCGCGCTAAAGATGTCTGGGTGACTTAGATGCTAGGATCAGTTGTTTCAATGTAAA	1192						
Qy	841	TGACACAGCCCGGACTCCGTACGGCACTAGCAGGGGACTGAAAGCGCTCTTCAGGTACTGC	900						
Db	1193	TGACACAGCCCGGACTCCGTACGGCACTAGCAGGGGACTGAAAGCGCTCTTCAGGTACTGC	1252						
Qy	901	TGTTGGGCGGTGATCGCTACAGGCGGATCAGACAGTTTGTCTTCTGGAACCTTGACA	960						
Db	1253	TGTTGGGCGGTGATCGCTACAGGCGGATCAGACAGTTTGTCTTCTGGAACCTTGACA	1312						
Qy	961	CTGCACACGGTAATGCTGA	980						
Db	1313	CTGCACACGGTAATGCTGA	1332						

RESULT 3
AC040169

LOCUS AC040169 212596 bp DNA linear PRI 17-DEC-2002
DEFINITION Homo sapiens chromosome 16 clone RP11-505K9, complete sequence.
ACCESSION AC040169
VERSION AC040169.7 GI:271511363
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 212596)
AUTHORS DOE Joint Genome Institute, Stanford Human Genome Center and Los Alamos National Laboratory.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 212596)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (11-APR-2000) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 212596)
AUTHORS DOE Joint Genome Institute, Stanford Human Genome Center and Los Alamos National Laboratory.
TITLE Direct Submission
JOURNAL Submitted (17-DEC-2002) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Dec 17, 2002 this sequence version replaced gi:18139274.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center and Los Alamos National Laboratory
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.9% of Sequence;
Estimated Total Number of Errors is 0.4.
NOTE: Shatter libraries failed to resolve dinucleotide repeat region from 8264 to 8629. Unsure number of repeat copies 8264 to 8629. Forced join at 8417.
Location/Qualifiers
1. 212596
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosomes="16"
/clone="RP11-505K9"
8264..8629
/note="NOTE: Shatter libraries failed to resolve dinucleotide repeat region from 8264 to 8629. Unsure number of repeat copies 8264 to 8629. Forced join at 8417."
misc_feature
ORIGIN
Query Match 98.4%; Score 964.8; DB 9; Length 212596;
Best Local Similarity 99.7%; Pred. No. 2.8e-193;
Matches 977; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
QY 1 CTCGCGGGAACAGCGCTGGGCACTCCATTGGGGCTGTTTACTCCCAACTCTCGCGAG 60
DB 173307 CTCGCGGGAACAGCGCTGGGCACTCCATTGGGGCTGTTTACTCCCAACTCTCGCGAG 173366
QY 61 ACTGGGCGGCTGGGCGAGCGAGCCACAGCTGGAGCTCAGCTCCGCGGACCCAGCGT 120
DB 173367 ACTGGGCGGCGGCGAGCGAGCCACAGCTGGAGCTCAGCTCCGCGGACCCAGCGT 173426
QY 121 GCCCTGTCTGCCCGCTCCCGGGCTTCGCTGGCGCTTCGAGCGCTGGGCGAGCGG 180
DB 173427 GCCCTGTCTGTCCCGCTCCCGGGCTTCGCTGGCGCTTCGAGCGCTGGGCGAGCGG 173486
QY 181 GACCACGCGGAGGATGACCAAGGTCTCCGCACTTTGCGCGCGGGGGCGCGGTGG 240
DB 173487 GACCACGCGGAGGATGACCAAGGTCTCCGCACTTTGCGCGCGGGGGCGCGGTGG 173545
QY 241 CAGGCTGGAAGCGGAGCGGTGGCGAGCTGCCAGCGCGCGGAGAACCGCTGGGGG 300
DB 173546 CAGGCTGGAAGCGGAGCGGTGGCGAGCTGCCAGCGCGGAGAACCGCTGGGGG 173605

QY 301 AACCTTGGTCCGCTCTGGCGCTCGCTCTAGGATCCCGAAAGAGGACGCGGCGCGAAA 360
DB 173606 AACCTTGGTCCGCTCTGGCGCTCGCTCTAGGATCCCGAAAGAGGACGCGGCGCGAAA 173665
QY 361 GCGGCAGGCTGGGCGAGGATCTAGAAAGACTGCTTGGCGCAGGCTCCCTGCCCGCGCGG 420
DB 173666 GCGGCAGGCTGGGCGAGGATCTAGAAAGACTGCTTGGCGCAGGCTCCCTGCCCGCGCGG 173725
QY 421 GCCTGCTGTCAAGGACTCGTGAGAGCTCGCTTCCGCGCGGAGCCCTTCTCTCAGGGGTC 480
DB 173726 GCCTGCTGTCAAGGACTCGTGAGAGCTCGCTTCCGCGCGGAGCCCTTCTCTCAGGGGTC 173785
QY 481 CACGTCCAGGCACCGCGGCTCGGCACACCCACCCCGCGCGGACCTGCCCTGGGTGC 540
DB 173786 CAGGTCCAGGCACCGCGGCTCGGCACACCCACCCCGCGCGGACCTGCCCTGGGTGC 173845
QY 541 CCTTTAAACCCGCGGCTAGCTCGTTAAGATGCGAAGTTCGCGTCCGGAACACGCGAAA 600
DB 173846 CCTTTAAACCCGCGGCTAGCTCGTTAAGATGCGAAGTTCGCGTCCGGAACACGCGAAA 173905
QY 601 CCCCAATCCCGCTCGCGACCTCTGACCCCGCGCGGCGGACGACGACTGGGCC 660
DB 173906 CCCCAATCCCGCTCGCGACCTCTGACCCCGCGCGGCGGACGACGACTGGGCC 173965
QY 661 TCCCGACGCGCAGCGCTGCGCGGACACCGGTGGGTGCGAAACGAGGACCTTTCTTAAC 720
DB 173966 TCCCGACGCGCAGCGCTGCGCGGACACCGGTGGGTGCGAAACGAGGACCTTTCTTAAC 174025
QY 721 GCCAGTGTGCTCTTTTGAAGAAACAAAGATTAATGTGTTAACTCTCTGAAAGCT 780
DB 174026 GCCAGTGTGCTCTTTTGAAGAAACAAAGATTAATGTGTTAACTCTCTGAAAGCT 174085
QY 781 TCCCGCTTAAAGATGCTCGGGTGTAGTCTAGGATCAGTTGTTTCAATGATAA 840
DB 174086 TCCCGCTTAAAGATGCTCGGGTGTAGTCTAGGATCAGTTGTTTCAATGATAA 174145
QY 841 TGGACACGCGCGGCTCCGTACGCGACTAGCAGGCGACTGAAAGCGCTCTTCAGGTACTGC 900
DB 174146 TGGACACGCGCGGCTCCGTACGCGACTAGCAGGCGACTGAAAGCGCTCTTCAGGTACTGC 174205
QY 901 TGGTGGGCGGTGATCGCTACAGGCGGATCAGACAGTTTGTCTCTGGAACCTTGACA 960
DB 174206 TGGTGGGCGGTGATCGCTACAGGCGGATCAGACAGTTTGTCTCTGGAACCTTGACA 174265
QY 961 CTGCACACGCGTAATGCTGA 980
DB 174266 CTGCACACGCGTAATGCTGA 174285

RESULT 4

AC139664/c

LOCUS

DEFINITION

AC139664

Homo sapiens clone RP11-812D13, WORKING DRAFT SEQUENCE, 9 unordered

pieces.

ACCESSION

AC139664

VERSION

AC139664.5

GT:29293970

HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

KEYWORDS

Homo sapiens (human)

SOURCE

Homo sapiens

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 212596)

Munz, D.M., Adams, C., Adio-Oduola, B., Ali-oman, F.R., Allen, C.,

Albrooks, S.L., Amarantunge, H.C., Are, J.R., Ayele, M., Banks, T.,

Barbaria, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D.,

Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,

Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,

Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,

Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,

Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,

Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,

Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,

Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
 Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
 Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
 Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
 Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B.,
 Homai, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
 Kratovich, J., Kureishi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louised, H.,
 Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwuonu, G.,
 Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
 Rivers, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G.,
 Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I.,
 Sodergren, B., Sonaite, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmami, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstock, G. and Gibbs, R.
 Direct Submission
 2 (bases 1 to 216522)
 Worley K.C.
 Direct Submission
 Submitted (08-FEB-2003) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 216522)
 Worley K.C.
 Direct Submission
 Submitted (27-MAR-2003) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Mar 27, 2003 this sequence version replaced gi:28467047.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Drafting Center Code: BCM
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: HET1
 Center clone name: RP11-812D13
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 215023 bases at least Q40
 Consensus quality: 215949 bases at least Q30
 Consensus quality: 216649 bases at least Q20
 Estimated insert size: 210467; sum-of-contigs estimation
 Quality coverage: 10x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 consists of 9 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence.
 * as soon as it is available and the accession number will

* be preserved.
 * 1 1056: contig of 1056 bp in length
 * 1057 1156: gap of unknown length
 * 1157 2298: contig of 1142 bp in length
 * 2299 2398: gap of unknown length
 * 2399 3437: contig of 1039 bp in length
 * 3438 3537: gap of unknown length
 * 3538 5421: contig of 1884 bp in length
 * 5422 5521: gap of unknown length
 * 5522 6548: contig of 1027 bp in length
 * 6549 7859: contig of 1211 bp in length
 * 7860 7959: gap of unknown length
 * 7960 21513: contig of 13554 bp in length
 * 21514 21613: gap of unknown length
 * 21614 40787: contig of 19174 bp in length
 * 40788 40887: gap of unknown length
 * 40888 216522: contig of 175635 bp in length.
 FEATURES
 source
 1. 216522
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clone="RP11-812D13"
 ORIGIN
 Query Match 98.4%; Score 964.8; DB 2; Length 216522;
 Best Local Similarity 99.7%; Pred. No. 2.8e-193;
 Matches 977; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
 QY 1 CTCGGCGGACACGCTGGGCACTCCATTCGGGGCTGTTTACTCCCAACTCTCGCGAG 60
 Db 19428 CTCGGCGGACACGCTGGGCACTCCATTCGGGGCTGTTTACTCCCAACTCTCGCGAG 19369
 QY 61 ACTGGCGCGCGGGCAGCGAGGCCACAGCTGGGAGCTTCAGCTCCGCGGACCCAGCGT 120
 Db 19368 ACTGGCGCGCGGGCAGCGAGGCCACAGCTGGGAGCTTCAGCTCCGCGGACCCAGCGT 19309
 QY 121 GCCCTGTCTGTCCCGCTCCCGGGCTTGGCTGGCGCTCTGAGCGCCGCTGGGAGCGG 180
 Db 19308 GCCCTGTCTGTCCCGCTCCCGGGCTTGGCTGGCGCTCTGAGCGCCGCTGGGAGCGG 19249
 QY 181 GACACGCGCGGAGATGCAAGAGTCTCGCGACATTCGCGGGCGGGGGCGCGTGG 240
 Db 19248 GACACGCGCGGAGATGCAAGAGT-CTCGGACATTCGCGGGCGGGGGCGCGTGG 19190
 QY 241 CAGGCTGGAAGCGGAGGGCGTGGCAGAGCTGCCAGCGCGGAGAACGCGCTGGGG 300
 Db 19189 CAGGCTGGAAGCGGAGGGCGTGGCAGAGCTGCCAGCGCGGAGAACGCGCTGGGG 19130
 QY 301 AACCCCTTGGTCCGCTCTGGCGCTCGCTTAGGATCCCGAAAGAGGACACGGGCGGAAA 360
 Db 19129 AACCCCTTGGTCCGCTCTGGCGCTCGCTTAGGATCCCGAAAGAGGACACGGGCGGAAA 19070
 QY 361 GCGCGCAGGCTGGCGCAGGATCTAGAAAGACTTCCTGGCGGAGCTTCCTGCCCCCGCG 420
 Db 19069 GCGCGCAGGCTGGCGCAGGATCTAGAAAGACTTCCTGGCGGAGCTTCCTGCCCCCGCG 19010
 QY 421 GCCTGCTGTCATGGACTCGTGAGAGCTCGCTTCCCGCGGAGACCTTCTCGCAGGGGTC 480
 Db 19009 GCCTGCTGTCATGGACTCGTGAGAGCTCGCTTCCCGCGGAGACCTTCTCGCAGGGGTC 18950
 QY 481 CAGCTCCAGGACCGCGGCTCGGACACCCACCCCGCGCGGACCTTGCCTTGGGTGC 540
 Db 18949 CAGCTCCAGGACCGCGGCTCGGACACCCACCCCGCGCGGACCTTGCCTTGGGTGC 18890
 QY 541 CCCTTAAACCGGGCGGTAGCTGTTAAGATGGGAGTGTCCGTCCGAAACACCGGAAA 600
 Db 18889 CCCTTAAACCGGGCGGTAGCTGTTAAGATGGGAGTGTCCGTCCGAAACACCGGAAA 18830
 QY 601 CCCCATAATCCCGCTGCCGACTCTCTGACCCCGCGGCCACGAGGACGACACTGGGCC 660
 Db 18829 CCCCATAATCCCGCTGCCGACTCTCTGACCCCGCGGCCACGAGGACGACACTGGGCC 18770

QY 661 TCCGACGCGCAGCGCGCTGCCGGGACACCGGTGCGTGCAGAAACGGAGGACCTTTGTAAAC 720
Db 18769 TCCGACGCGCAGCGCGCTGCCGGGACACCGGTGCGTGCAGAAACGGAGGACCTTTGTAAAC 18710
QY 721 GCCACGTGTTGCTCTTTTGAAGAAACAGAAATAAATGTGTTAAACTGCTGAAAGCT 780
Db 18709 GCCACGTGTTGCTCTTTTGAAGAAACAGAAATAAATGTGTTAAACTGCTGAAAGCT 18650
QY 781 TGCCGGCTAAAGACTGCTGGGTGACTTAGACTCTAGGATCAGTTGTTTCAATGTA 840
Db 18649 TGCCGGCTAAAGACTGCTGGGTGACTTAGACTCTAGGATCAGTTGTTTCAATGTA 18590
QY 841 TGACACGCGCGGACTCCGTAACGCGACTAGCAGGAGCTGAAACGCTCTTCAGGTACTGC 900
Db 18589 TGACACGCGCGGACTCCGTAACGCGACTAGCAGGAGCTGAAACGCTCTTCAGGTACTGC 18530
QY 901 TGGTGGGGCGGTGATGCGGTACAGGCGGATCAGACAGTTTGTGCTCTGGAACCTTGACA 960
Db 18529 TGGTGGGGCGGTGATGCGGTACAGGCGGATCAGACAGTTTGTGCTCTTGGAACTTGACA 18470
QY 961 CTGCACACCGGTAATGCTGA 980
Db 18469 CTGCACACCGGTAATGCTGA 18450

BC052334 532 bp mRNA linear PRI 05-MAY-2003
LOCUS Homo sapiens cDNA clone IMAGE:6176570, partial cds.
DEFINITION BC052334
ACCESSION BC052334
VERSION BC052334.1 GI:30354085

KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Uedin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullaly, S.J., Bosak, S.A., McSwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahy, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalls, D.E.,
Scherch, A., Schein, J.E., Jones, S.J. and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22388257
12477932
2 (bases 1 to 532)
Strausberg, R.
Direct Submission
Submitted (02-MAY-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabbs@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.

TITLE

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
JOURNAL

REMARK
COMMENT

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>

Contact: nisc_mgc@nih.gov
Akhter, N., Ayale, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,
Maduro, Q.L., Mastello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,
Tsurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 115 Row: n Column: 1
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, GenomeScan gene prediction.

FEATURES

source

1. 532
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6176570"
/tissue_type="uterus, leiomyosarcoma"
/clone_lib="NIH_MGC_71"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"

ORIGIN

Query Match 53.2%; Score 521.4; DB 9; Length 532;
Best Local Similarity 99.8%; Pred. No. 1.3e-99;
Matches 522; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 272 GCTGCCAGGCGCGAGAACGCGCTGGGGAAACCCCTTGGTCCGCTCTGCGCGTCCGCTCTAG 331
Db 1 GCTGCCAGGCGCGAGAACGCGCTGGGGAAACCCCTTGGTCCGCTCTGCGCGTCCGCTCTAG 60
QY 332 GATCCCCGAAAAAGGAGCAGCGGCGGAAACGCGGCCAGGCTGGGCCAGGATCTAGAAAGAC 391
Db 61 GATCCCCGAAAAAGGAGCAGCGGCGGAAACGCGGCCAGGCTGGGCCAGGATCTAGAAAGAC 120
QY 392 TGCCTGGGCGCAGCTCCCTGCCCCCGCGGCTCTGCTCATGAGCTCTGAGAGCTCGC 451
Db 121 TGCCTGGGCGCAGCTCCCTGCCCCCGCGGCTCTGCTCATGAGCTCTGAGAGCTCGC 180
QY 452 TTCCCGCGCGGACCCCTTCTCTGAGGGGTCCACGCTCCAGGACCGCGGCGCTCGGACACCCC 511
Db 181 TTCCCGCGCGGACCCCTTCTCTGAGGGGTCCACGCTCCAGGACCGCGGCGCTCGGACACCCC 240
QY 512 ACCCCCGCGCGGCGACCTTCCCTCGGTGTCCTTAAACCCGCGGCTAGCTCGTTAAAGATG 571
Db 241 ACCCCCGCGCGGCGACCTTCCCTCGGTGTCCTTAAACCCGCGGCTAGCTCGTTAAAGATG 300
QY 572 GCGAAGTGTCCGCTCGGAAACGCGGAAACCCCAATCCGCTGCCCCACCTCTCGACC 631
Db 301 GCGAAGTGTCCGCTCGGAAACGCGGAAACCCCAATCCGCTGCCCCACCTCTCGACC 360
QY 632 CCGCGCCCCACGCGGACGACAGACTCGGCTCCCGCGCGCAGCGCGCTCCCGGACACCG 691
Db 361 CCGCGCCCCACGCGGACGACAGACTCGGCTCCCGCGCGCAGCGCGCTCCCGGACACCG 420
QY 692 GTGCGTGGCAACGCGGAGGACCTTTGTAAACGCCACGCTGTTGTCTTTTGAACAAAG 751
Db 421 GTGCGTGGCAACGCGGAGGACCTTTGTAAACGCCACGCTGTTGTCTTTTGAACAAAG 480
QY 752 AATAAATGTGTTAAACTGCTGAAAGCTTGGCGCGCTAAAGAG 794
Db 481 AATAAATGTGTTAAACTGCTGAAAGCTTGGCGCGCTAAAGAG 523

RESULT 6
CQ729132/c
LOCUS CQ729132 4338 bp DNA linear PAT 03-FEB-2004
DEFINITION Sequence 15066 from Patent WO02068579.
ACCESSION CQ729132
VERSION CQ729132.1 GI:42299883
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1
Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.
Kits, such as nucleic acid arrays, comprising a majority of
humanexons or transcripts, for detecting expression and other uses
thereof
JOURNAL Patent: WO 02068579-A 15066 06-SEP-2002;
PE Corporation (NY) (US)
FEATURES Location/Qualifiers
source 1. 4338
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 12.9%; Score 126; DB 6; Length 4338;
Best Local Similarity 100.0%; Pred. No. 3.1e-16;
Matches 126; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCGGCGGCAACACGCGCTGGCACTCCATTGGGGCTGTTTACTCCAACTCTCGCGAG 60
Db 126 CTCGGCGGCAACACGCGCTGGCACTCCATTGGGGCTGTTTACTCCAACTCTCGCGAG 67

QY 61 ACTGGGGCGGGCGGCCACGAGGCCACAGCTGGGAGCCTCAGCTCGCGCCGACCCAGCGT 120
Pb 66 ACTGGGGCGGGCGGCCACGAGGCCACAGCTGGGAGCCTCAGCTCGCGCCGACCCAGCGT 7

QY 121 GCCCTG 126
Db 6 GCCCTG 1

RESULT 7
BD244989/c
LOCUS BD244989 4338 bp DNA linear PAT 17-JUL-2003
DEFINITION Mammalian subtilisin/kexin isoenzyme SKI-1: Protein convertase
having peculiar cleaving characteristic.
ACCESSION BD244989
VERSION BD244989.1 GI:33054759
KEYWORDS JP 2002532065-A/3.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4338)
AUTHORS Seidan, N., Chretien, M., Marcinkiewicz, M., Laaksonen, R. and
Davignon, J.
TITLE Mammalian subtilisin/kexin isoenzyme SKI-1: Protein convertase
having peculiar cleaving characteristic
JOURNAL Patent: JP 2002532065-A 3 02-OCT-2002;
INSTITUT DE RECHERCHES CLINIQUES DE MONTREAL
COMMENT OS Homo sapiens (human)
PN JP 2002532065-A/3
PD 02-OCT-2002
PF 04-NOV-1999 JP 2000579720
PR 04-NOV-1998 CA 2249648
PI NABIL SEIDAH, MICHEL CHRETIEN, MIECZYSLAW MARCINKIEWICZ, RELJO
PI LAAKSONEN, R.
PI JEAN DAVIGNON
PC C12N1/15/09, A61K38/00, A61K45/00, A61P1/16, A61P3/06, A61P9/10, PC
A61P25/28.
PC A61P35/00, A61P43/00, C07K7/06, C07K7/08, C12N1/15, C12N1/19, C12N1/21, C12N1/28.


```
Db          6  GCCTG 1

RESULT 9
AR527529/c
LOCUS      AR527529
DEFINITION Sequence 3 from patent US 672353.
ACCESSION  AR527529
VERSION     AR527529.1 GI:53914634
KEYWORDS   Unknown.
SOURCE      Unknown.
ORGANISM    Unclassified.
REFERENCE   1  (bases 1 to 4338)
AUTHORS     Abrahamson, L., Ekblom, J., Forsgren, M., Horling, J. and Johansson, P.
TITLE       Promoter sequences
JOURNAL     Patent: US 672353-A 3 20-APR-2004;
FEATURES    Location/Qualifiers
            source
            1..4338
            /organism="unknown"
            /mol_type="genomic DNA"

ORIGIN
Query Match      12.7%; Score 124.4; DB 6; Length 4338;
Best Local Similarity 99.2%; Pred. No. 6.7e-16;
Matches 125; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1  CTCGGGGGACACAGCCCTGGGCACCTCCATTGGGGCTGTTTACTCCCAACTCTCGCGAG 60
        |||||||
Db      126 CTCGGGGGACACAGCCCTGGGCACCTCCATTGGGGCTGTTTACTCCCAACTCTCGCGAG 67
        |||||||

QY      61  ACTGGGGCGGGCGGCGAGCGAGCCACACAGCTGGGAGCCTCAGCTCCGCGGAGCCAGCGT 120
        |||||||
Db      66  ACTGGGGCGGGCGGCGAGCGAGCCACACAGCTGGGAGCCTCAGCTCCGCGGAGCCAGCGT 7
        |||||||

QY      121 GCCTG 126
        |||||||
Db      6  GCCTG 1

RESULT 11
HUMKIAAL/c
LOCUS      HUMKIAAL
DEFINITION Homo sapiens KIAA0091 mRNA, partial cds.
ACCESSION  D42053
VERSION     D42053.1 GI:577308
KEYWORDS   Homo sapiens (human)
SOURCE      Homo sapiens
ORGANISM    Homo sapiens
REFERENCE   1
AUTHORS     Nagase, T., Miyajima, N., Tanaka, A., Suzuki, T., Seki, N., Sato, S.,
            Tabata, S., Ishikawa, K.-I., Kawarabayashi, Y., Kotani, H. and Nomura, N.
            Prediction of the coding sequences of unidentified human genes.
            III. The coding sequences of 40 new genes (KIAA0081-KIAA0120)
            deduced by analysis of cDNA clones from human cell line KG-1
JOURNAL     DNA Res. 2 (1), 37-43 (1995)
MEDLINE     95308325
PUBMED      7788527
REFERENCE   2  (bases 1 to 4338)
AUTHORS     Ohara, O., Nagase, T., Kikuno, R. and Nomura, N.
TITLE       Direct Submission
JOURNAL     Submitted (08-NOV-1994) Osamu Ohara, Kazusa DNA Research Institute;
            1532-3, Yana, Kisarazu, Chiba 292-0812, Japan
            (E-mail: cdnainfo@kazusa.or.jp, Tel: +81-438-52-3913)
FEATURES    Location/Qualifiers
            1..4338
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /sex="male"
            /cell_line="KG-1"
            /cell_type="myeloblast"
            1..4338
            /gene="KIAA0091"
            1..496
            /gene="KIAA0091"
            <479..3655
            CDS

gene
5' UTR
CDS

RESULT 10
AX024613/c
LOCUS      AX024613
DEFINITION Sequence 5 from Patent WO0026348.
ACCESSION  AX024613
VERSION     AX024613.1 GI:10184749
KEYWORDS   Homo sapiens (human)
SOURCE      Homo sapiens
ORGANISM    Homo sapiens
REFERENCE   1
AUTHORS     Davignon, J., Chretien, M., Laaksonen, R., Seidah, N. and
            Marcinkiewicz, M.
TITLE       Mammalian subtilisin/kexin isozyme ski-1: a proprotein
JOURNAL     Patent: WO 0026348-A 5 11-MAY-2000;
            DAVIGNON JEAN (CA); CHRETIEN MICHEL (CA); LAAKSONEN RELJO (CA);
            SEIDAH NABILA (CA); MARCINKIEWICZ MIECZYSLAW (CA); MONTREAL INST
            RECH CLINIQUES (CA)
FEATURES    Location/Qualifiers
            1..4338
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"
            /db_xref="taxon:9606"
            497..3655
            /note="unnamed protein product"
            /codon_start=1
            /protein_id="CAC08995.1"
            /db_xref="GI:10184750"
            /translation="MKLVNIWLLLVLLVLCGKKHGLDRLEKSKFKAFCPCGSGSHLTLK
            VFESTVVEYIVAFNGYFTAKRNSFISSALKSEVDNWRRIIPNNPSSDYPDFE
            VTQIKKQAGLTLLEDHNIKRVTPQRKFRSLKYAESDTPVPCNETRWSQWSSR
            PLRRASLSIGSGFWHATGHRSSRRLLRAIPRQVAQTLDVLMQNGYTGANVRVAVFD"
            CDS
```

```

/gene="KIAA0091"
/Note="KIAA0091 gene product is related to subtilisin.
Start codon is not identified"
/citation=[1]
/codon_start=1
/protein_id="BAA07653.2"
/db_xref="GI:40789961"
/translation="SYSLVTMKLVNIMLLVLLCGKKHGLDRLEKSPKAPCPG
SHLTKEVSTVVEYEVAFNGYFTAKARNSFISSALKSEVDNWRILIRNPSSD
YPSDFEVQIKKQAGLLTLEDHNTKRVTPQRKSLKIAESDTPVPCNETRWSQ
KWQSRPLRRASLSLGSFHWATGRHSRLRLRAIPRQAQTLQADVLQWQYTGANV
RVAVFDTLGSKHPFKNKTNTNERTLDGLGHTFVAGVIASRQCQGFAPDA
ELHVPRTNKQSVTFSLDAFNVAIKKIDVNLISIGDPDMHDPFVKWELTAN
NVIWUSALGNQDPLGLNLPADQMDVIGVGLDPEDNIARFSSRGMNTWELPGYGR
MKPDIVTVAGVGRSGVGRGCRALSGTSVASPVVAGATVLLVSTVQKRLVNPASMKQ
ALIASARVNMFEQGHGKLDLRLAYOILNSYKQASLSPSYIDLTECPYMPYCS
OPIYVGMPTVNVTVILNGMGVTRIVDKPDQWQYLPQNGDNIIEVSYSLVPMWSG
YLATISVTKKAASWEGIAQGHVMTIVASPAETESKGAETSTVKLPKIKIPTTP
RSKVLMDQVHNLRYPPGVFPRDNLRMKNDPLDWDGDIHTNFRDMYOHLSRMSGVFE
VLGAFTCFDASQYGLTLMVDSBEYFPEEIAKLRRDNDGLSLVIFSDWYNTSMRK
VKFYDENTRQWMMPTDGTGANIPALNELLISVWNNPFDGLYEGEFTLANHDMYASGCS
IAKPEDGVITQPDQGLVYLKQETAVENPILGYQIPAEQGGRIYVIGDSNCL
DHSRQKOCFLWLLDQVTSYVTPPSLSHSGNRQRPSPGAGSVTPERMENGLHRY
SKVLRAHLGDPKPRPLPACPLRSWAKPQPLNETAPSNLWKQKLSIDLKVLPLNFR
SNRPQVRPLSPGESAWDIPGIMPRYQVQGTIPVPAFLGAMVVLAFVQVQINKA
KSRPKRRPRVPRKPOLMQVHPKTPSV"
3656..4338
/gene="KIAA0091"

```

ORIGIN

```

Query Match 12.7%; Score 124.4; DB 9; Length 4338;
Best Local Similarity 99.2%; Pred. No. 6.7e-16;
Matches 125; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTCGGCGGGAACAGCCCTGGGCACTCCACTTCGGGCTGTTTACTCCCAACTCTCGCGAG 60
pb 126 CTCGGCGGGAACAGCCCTGGGCACTCCACTTCGGGCTGTTTACTCCCAACTCTCGCGAG 67

Qy 61 ACTGGCGGCGCGGCGGAGCGAGCCACAGCTGGGAGCCTCAGCTCCGCGGAGCCAGCGT 120
Db 66 ACTGGCGGAGCGGCGGAGCGGAGCCACAGCTGGGAGCCTCAGCTCCGCGGAGCCAGCGT 7

Qy 121 GGCCTG 126
Db 6 GGCCTG 1

```

```

RESULT 12
LOCUS BC026330/c
DEFINITION Homo sapiens membrane-bound transcription factor protease, site 1,
ACCESSION BC026330
VERSION BC026330.1 GI:20072867
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3026)
AUTHORS Strausberg,R.D., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Schaeetz,T.E., Brownstein,M.J., Ustin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullaly,S.J., Bosak,S.A., McSwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Workley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodrigues,S.,

```

Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalls,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

1247932

2 (bases 1 to 3026)

Strausberg,R. Direct Submission

Submitted (02-APR-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgapbs-remail.nih.gov

Tissue Procurement: Miklos Palkovits, M.D., Ph.D.

CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshiyuki and Piero Carninci (RIKEN)

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305

Web site: <http://www-shgc.stanford.edu>

Contact: (Dickson, Mark) mcd@paxil.stanford.edu

Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAC Plate: 32 Row: m Column: 11

This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4506774

This clone has the following problem: The cds is short compared to the longest cds in the locus.

Location/Qualifiers

1..3026

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:4816102"

/tissue_type="Brain, hippocampus"

/clone_lib="NIH_MGC_95"

/lab_host="DH10B"

/note="Vector: pBluescript"

1..3026

/gene="MBTPS1"

/note="Synonyms: SLP, SKI-1, KIAA0091, PCSK8"

/db_xref="LocusID:8720"

/db_xref="MIM:603355"

463..2121

/gene="MBTPS1"

/codon_start=1

/product="MBTPS1 protein"

/protein_id="AAH26330.1"

/db_xref="GI:20072868"

/db_xref="LocusID:8720"

/db_xref="MIM:603355"

/translation="MKLVNIMLLVLLCGKKHGLDRLEKSPKAPCPGCSHLTLK VEFSTVVEYEVAFNGYFTAKARNSFISSALKSEVDNWRILIRNPSSDYPSPDFE YIQIKKQAGLLTLEDHNTKRVTPQRKSLKIAESDTPVPCNETRWSQKWQSR PLRRASLSLGSFHWATGRHSRLRLRAIPRQAQTLQADVLQWQYTGANVAVFD TGLSKHPFKNKTNTNERTLDGLGHTFVAGVIASRQCQGFAPDAELHIFR VFTNQSVASWFLDAFNVAIKKIDVNLISIGDPDMHDPFVKWELTANNVIVS AIGNQDPLGLNLPADQMDVIGVGLDPEDNIARFSSRGMNTWELPGYGRMKPDIV TYGAVGRSGVGRGCRALSGTSVASPVVAGATVLLVSTVQKRLVNPASMKQALIASA RRLPGVNMFEQGHGKLDLRLAYOILNSYKQASLSPSYIDLTECPYMPYCSQPIYYG 1045..1749

/gene="MBTPS1"

/note="Peptidase_S8; Region: Subtilase family. Subtilases

misc_feature


```

/gene="JPH3"
/notes="component of the junctional complex between plasma
membrane and endoplasmic reticulum"
/codon_start=1
/product="junctophilin 3"
/db_xref="GI:17646245"
/translation="MSSGGRFNFDDGSGYCGWEDGKAHGHGVCCTGPKGQGEYTGSGWS
HGFEVLGVYTPSGNTYGTWAGQKRHIGLESKGKVVYKGSWTHGPKRGYVRECA
NGAKYEGTWSNQLQDGYGTETYSDD"

ORIGIN
Query Match          6.6%; Score 64.2; DB 9; Length 125020;
Best Local Similarity 11.6%; Pred. No. 0.0029;
Matches 61; Conservative 260; Mismatches 202; Indels 2; Gaps 1;

QY 44 CTCCTCACTCTCCGAGACTGGCGCGCGCGGCGGAGGAGCCAGCTGGAGGCTCAG 103
Db 50678 VDVVBRWBDBTSWGSBHYMRGVHTGVVRWVGYGBDTHVYDGGHSGWMKSCGYSD 50737

QY 104 CTCGCGGACCCAGCGTCCCTGTCTCCCGCGCTCCGCGGCTTGGCTGGCGCTCTG 163
Db 50738 BGKCAKBSKMTCTGSMKTKTCYGYKWSKADBSYYSBHYBWSYMCCKRBSYASS 50797

QY 164 GACGCGTGGCGAGCGGACACGCGCGGAGGATGACGAAGTGCTCCGACATTTGG 223
Db 50798 YDVRYKBTGMRYVRSWYVYKBSMSVSARKSGHKVRSBMSKCSRRKDMSTSSMWR 50857

QY 224 GCGCGCGCGCGCGTGGCAGGTGGAAGCGGAGGCGGTGGCGAGCTCCAGCGCG 283
Db 50858 SMKDRKCSRSHSYKSKMGKVTSDVDRCYACSBMKYKMRHKKHKKDHSBWSGKSHMK 50917

QY 284 CGAGAACGCGCTGGGGAACCTTGTCGCTGCGCTGCTGCTAGGATCCCGGAAA 343
Db 50918 KGAWRVRSYMRSCSVHDSWVDSRRRMRKGSAMRGMKCYTSSMRSTKSRMSKSWR 50977

QY 344 GGAGCACGCGCGCGAAGCGGCGGAGGATCTAGAGACTAGAGACTGCTGGCGCG 403
Db 50978 WSGKCYCYGYVWSCTKMRSMGCSYSTGSGSMKMGKYCWGAGRSRYCKSGSRMSMAG 51037

QY 404 GCTCCCTGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 463
Db 51038 SMSASMSMGKMTSGRGRSARMSYCSRCRCKSKRCSCGCMRCWYCWYKCMRYGSM 51097

QY 464 CCTTCTCAGCGGTCACGTCACGCGCGCGCTGCGGCGCGCTGCGGCGCGCGCG 523
Db 51098 RMYKCCMSYKTKYKGCSEA--TKSMAYYSCHMRWRKGTSSAWWYMSMCCSKARAASGWY 51155

QY 524 GCACCTGCGCTGGTGGCTCCCTTAACCGCGGCGGTAGCTGCTGTTAAG 568
Db 51156 KCMHMKGGCYGGGKGSASTWYMMCCSRACKYTYTTRKSSRGRWG 51200

RESULT 15
AC079431
LOCUS Mus musculus chromosome 16 clone RP23-1J12, WORKING DRAFT SEQUENCE,
DEFINITION 77 unordered pieces.
AC079431
AC079431.1 GI:9958043
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 300695)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Mouse
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 300695)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (01-SEP-2000) Production Sequencing Facility, DOE Joint

```

COMMENT

Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 -----Genome Center
 Center: Joint Genome Institute
 Center Code: JGI
 Web site: http://www.jgi.doe.gov

 Project Information
 Center Project Name: 0
 Center clone name: RPCI-23_1J12

 Summary Statistics
 Consensus quality: 164936 bases at least Q40
 Consensus quality: 197223 bases at least Q30
 Consensus quality: 214345 bases at least Q20
 Estimated insert size: 257300; agarose-fp estimation
 Estimated insert size: 293095; sum-of-contigs estimation
 Quality coverage: 3.01 in Q20 bases; agarose-fp estimation
 Quality coverage: 2.65 in Q20 bases; sum-of-contigs estimation.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 77 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1076: contig of 1076 bp in length
 1077 1176: gap of unknown length
 1177 2208: contig of 1032 bp in length
 2209 2308: gap of unknown length
 2309 3527: contig of 1219 bp in length
 3528 3627: gap of unknown length
 3628 4764: contig of 1137 bp in length
 4765 4864: gap of unknown length
 4865 5952: contig of 1088 bp in length
 5953 6052: gap of unknown length
 6053 7182: contig of 1130 bp in length
 7183 7283: gap of unknown length
 7284 8420: contig of 1138 bp in length
 8421 8520: gap of unknown length
 8521 9616: contig of 1096 bp in length
 9617 9717: gap of unknown length
 9718 11229: contig of 1413 bp in length
 11230 12408: contig of 1179 bp in length
 12409 12508: gap of unknown length
 12509 13767: contig of 1259 bp in length
 13768 13867: gap of unknown length
 13868 15044: contig of 1177 bp in length
 15045 15144: gap of unknown length
 15145 16325: contig of 1181 bp in length
 16326 16425: gap of unknown length
 16426 17578: contig of 1153 bp in length
 17579 17679: gap of unknown length
 17680 18845: contig of 1167 bp in length
 18846 18946: gap of unknown length
 18947 20061: contig of 1115 bp in length
 20062 20160: gap of unknown length
 20161 21258: contig of 1098 bp in length
 21259 21358: gap of unknown length
 21359 22390: contig of 1032 bp in length
 22391 23599: contig of 1109 bp in length
 23600 23699: gap of unknown length
 23700 24907: contig of 1208 bp in length
 24908 25007: gap of unknown length
 25008 26211: contig of 1114 bp in length
 26212 26222: gap of unknown length
 26223 27273: contig of 1052 bp in length
 27274 27373: gap of unknown length
 27374 28472: contig of 1099 bp in length
 28473 28572: gap of unknown length
 28573 29683: contig of 1111 bp in length

Qy 643 GGGACGACAG 652
Db 13437 CCGNNGCCG 13446

Search completed: October 9, 2005, 02:21:59
Job time : 4302 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 11, 2005, 22:51:27 ; Search time 7895 Seconds
(without alignments)

11903.823 Million cell updates/sec

Title: US-10-829-118-2

Perfect score: 2469

Sequence: 1 ttgagctctctggaggtc.....gaatacatggcctacagct 2469

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST:*

1: gb_est1:.*
2: gb_est2:.*
3: gb_hic:.*
4: gb_est3:.*
5: gb_est4:.*
6: gb_est5:.*
7: gb_est6:.*
8: gb_gss1:.*
9: gb_gss2:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	471.6	19.1	500	6	CB049709 NISC_gj12
C 2	468	19.0	468	1	AI792242 ov05e09.y
C 3	451.8	18.3	455	1	AI760651 w166g09.x
C 4	448	18.1	450	2	AW051661 wx27b06.x
C 5	440.6	17.8	1240	5	BQ437878 AGENCOURT
C 6	435	17.6	435	1	AI685715 tu37c07.x
C 7	435	17.6	435	1	AI733521 ov05e09.x
C 8	435	17.6	435	2	BF446188 7p33f05.x
C 9	432	17.5	432	1	AI554846 te45b04.x
C 10	430.2	17.4	436	2	AW005919 w290e09.x
C 11	411.8	16.7	415	1	AI362974 qy87h07.x
C 12	404	16.4	404	2	BE048839 hr54e04.x
C 13	362.8	14.7	389	1	AI498080 tm66a04.x
C 14	332	13.4	3491	3	BC039100 Homo sapi
C 15	305.2	12.4	897	6	CD557847 AGENCOURT
C 16	303.6	12.3	457	2	BF570483 602186293
C 17	299.6	12.1	15970	8	AQ839852 260113-C5
C 18	288.4	12.1	881	5	BU521286 AGENCOURT
C 19	280.4	11.8	736	6	CA427039 UI-H-DF0
C 20	288.2	11.7	2786	3	BC035101 Homo sapi
C 21	287.6	11.6	918	5	BQ706343 AGENCOURT
C 22	282.4	11.4	672	6	CA431692 UI-H-DF0
C 23	280.8	11.4	3180	3	HSM800699
C 24	280.6	11.4	2097	3	CR614786 full-leng

25	279.4	11.3	630	5	BX508391
26	277.4	11.2	629	5	BX509360
27	277.2	11.2	778	1	AV764490
C 28	276	11.2	3474	3	HSM800886
C 29	274.8	11.1	834	8	BZ611716 WHAA044TR
C 30	273.8	11.1	1027	5	BQ70852
C 31	273.8	11.1	1036	5	BQ070834
C 32	273.8	11.1	1036	5	BQ070834
C 33	273.6	11.1	2558	3	BC038553
C 34	273.4	11.1	3016	3	BC068461
C 35	273.2	11.1	4828	3	HSM802759
C 36	272.6	11.0	5479	3	HSM807678
C 37	272	11.0	610	7	CR540737
C 38	271.8	11.0	923	6	CD518127
C 39	271	11.0	910	5	BQ722917
C 40	270.8	11.0	3840	3	BC032939
C 41	270	10.9	3566	3	HSM805040
C 42	269.8	10.9	694	7	CR764988
C 43	269.2	10.9	680	5	AQ356404
C 44	268.8	10.9	660	5	BX484570
C 45	268.8	10.9	769	5	BQ710315

ALIGNMENTS

RESULT 1
CB049709/c

LOCUS NISC_gj12f01.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:3271441
DEFINITION 3', mRNA sequence.

ACCESSION CB049709

VERSION CB049709.1 GI:27787996

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 500)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov

CDNA Library Preparation:
CDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC)

Cloned through the I.M.A.G.E. Consortium/LLNL at:
found through the I.M.A.G.E. Consortium/LLNL at:
Infoimage.llnl.gov

Plate: LLAM8008 row: L column: 2

Seq primer: -21M13 forward primer (ABI).

FEATURES
1. 500
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3271441"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Pr28"

/notes="Organ: prostate; Vector: pT73D-Pac (Pharmacia)
with a modified polylinker; Plasmid DNA from the
normalized library NCI CGAP Pr22 was prepared, and ss
circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (clonoids
985608-986759, 1101192-1101959, and 1217928-1220615).
Subtraction by Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 19.1%; Score 471.6; DB 6; Length 500;
 Best Local Similarity 99.2%; Pred. No. 6.5e-66;
 Matches 474; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 652 GAACCTTGGTCGCTCTGCGCTGCTCTAGGATCCCGAAAGAGGACGCGCGCGAA 711
 |||||
 Db 495 GAGGCTTGGTCGCTCTGCGCTGCTCTAGGATCCCGAAAGAGGACGCGCGCGAA 436
 |||||

QY 712 AGCGCCAGCTGGGCGAGATCTAGAAAGACTGCTGGGCGAGGCTCCTGCGCCCGCG 771
 |||||
 Db 435 AGCGCCAGCTGGGCGAGATCTAGAAAGACTGCTGGGCGAGGCTCCTGCGCCCGCG 376
 |||||

QY 772 GGCTGTGTCATGAGCTGTGAGAGCTGCTTCCGCGCGACCTTCTCTCAGGGGT 831
 |||||
 Db 375 GGCTGTGTCATGAGCTGTGAGAGCTGCTTCCGCGCGACCTTCTCTCAGGGGT 316
 |||||

QY 832 CCAGTCCAGGACCGGCGCTCGGACACCCACCCCGCGCGGACCTGCGCTCGGGTG 891
 |||||
 Db 315 CCAGTCCAGGACCGGCGCTCGGACACCCACCCCGCGCGGACCTGCGCTCGGGTG 256
 |||||

QY 892 CCCTTAAACCGGCGGTAGCTGTTAAGATGGCGAAGTGTCCGGTCCGGAAACACGCGAA 951
 |||||
 Db 255 CCCTTAAACCGGCGGTAGCTGTTAAGATGGCGAAGTGTCCGGTCCGGAAACACGCGAA 196
 |||||

QY 952 ACCGCAATCCCGCTGCCGACTCTGACCCCGCGCGCGGACGACGACTGGGC 1011
 |||||
 Db 195 ACCGCAATCCCGCTGCCGACTCTGACCCCGCGCGCGGACGACGACTGGGC 136
 |||||

QY 1012 CTCCGAGCGGACGCGCTGCGGAGACACCGGTGGTGGGAAACGAGGACCTTTGTAA 1071
 |||||
 Db 135 CTCCGAGCGGACGCGCTGCGGAGACACCGGTGGTGGGAAACGAGGACCTTTGTAA 76
 |||||

QY 1072 CGCCAGTGTTCCTCTTTTGAAGAAACAAAGAAATAAATGTGTTAACTGTCTGAAA 1129
 |||||
 Db 75 CGCCAGTGTTCCTCTTTTGAAGAAACAAAGAAATAAATGTGTTAACTGTCTCAAAAA 18
 |||||

RESULT 2
 AI792242
 LOCUS
 DEFINITION ov05e09.y5 NCI_CGAP_kid3 Homo sapiens cDNA clone IMAGE:1636456 5', mRNA linear EST 02-JUL-1999
 mRNA sequence.

ACCESSION AI792242 GI:53399949
 VERSION
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCI/NIH-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute / National Institute of Dental Research,
 Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 Unpublished (1997)
 JOURNAL Other_Ests: ov05e09.x5
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-i@mail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www-bio.llnl.gov/bbrp/image/image.html

This read is a RESEQUENCE of a previously sequenced human clone
 Original clone citation: see original entry for original citation
 Information
 This 5' resequenced clone has no previous 5' data to verify this
 new read against

Putative full length read
 The vector to vector length is 469
 Seq primer: -40RP from Gibco.

FEATURES
 source
 1..468
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:1636456"
 /lab_host="DH10B"
 /clone_lib="NCI_CGAP_Kid3"
 /note="Organ: kidney; Vector: p773D-Pac (Pharmacia) with
 a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
 strand cDNA was primed with a Not I - oligo(dT) primer,
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not
 I and Eco RI sites of the modified p773 vector. mRNA
 source: 2 pooled kidneys. Library went through one round
 of normalization. Library constructed by Bento Soares and
 M. Fatima Bonaldo."

ORIGIN

Query Match 19.0%; Score 468; DB 1; Length 468;
 Best Local Similarity 100.0%; Pred. No. 2.5e-65;
 Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 656 CTTTGGTCGCTCTGCGCTGCTCTAGGATCCCGAAAGAGGACGCGCGCGAAAGCG 715
 |||||
 Db 1 CTTTGGTCGCTCTGCGCTGCTCTAGGATCCCGAAAGAGGACGCGCGCGAAAGCG 60
 |||||

QY 716 GCAGGCTGGGCGCAGGATCTAGAAAGACTGCTTGGCGCAGGCTCCTGCGCCCGCGGGCC 775
 |||||
 Db 61 GCAGGCTGGGCGCAGGATCTAGAAAGACTGCTTGGCGCAGGCTCCTGCGCCCGCGGGCC 120
 |||||

QY 776 TGTCTGATGAGTCTGTGGAGAGCTGCTTCCGCGCGGACCTTCTCTGAGGGGTCCAC 835
 |||||
 Db 121 TGTCTGATGAGTCTGTGGAGAGCTGCTTCCGCGCGGACCTTCTCTGAGGGGTCCAC 180
 |||||

QY 836 GTCCAGGACCGCGGCTCGGACACCCACCCCGCGCGGACCTGCGCTGGGTGCCCC 895
 |||||
 Db 181 GTCCAGGACCGCGGCTCGGACACCCACCCCGCGCGGACCTGCGCTGGGTGCCCC 240
 |||||

QY 896 TTAAACCCCGCGGTAGCTGCTTAAAGATGCGAAGTGTCCGGTCCGGAAACACGCGAAACCC 955
 |||||
 Db 241 TTAAACCCCGCGGTAGCTGCTTAAAGATGCGAAGTGTCCGGTCCGGAAACACGCGAAACCC 300
 |||||

QY 956 CAAATCCCGCTTCCCGACCTCTGACCCCGCGCGGACGACAGACTGGGCTTCC 1015
 |||||
 Db 301 CAAATCCCGCTTCCCGACCTCTGACCCCGCGCGGACGACAGACTGGGCTTCC 360
 |||||

QY 1016 CGACGCGGAGCGGCTGCGGAGACACCGTGGTGGGAAACGAGGACCTTTGTACGCGC 1075
 |||||
 Db 361 CGACGCGGAGCGGCTGCGGAGACACCGTGGTGGGAAACGAGGACCTTTGTACGCGC 420
 |||||

QY 1076 ACGTGTTCCTCTTTTGAAGAAACAAAGAAATAAATGTGTTAACTGTC 1123
 |||||
 Db 421 ACGTGTTCCTCTTTTGAAGAAACAAAGAAATAAATGTGTTAACTGTC 468
 |||||

RESULT 3
 AI760651/c
 LOCUS
 DEFINITION w16909.x1 NCI_CGAP_Kid12 Homo sapiens cDNA clone IMAGE:2398336 3', mRNA linear EST 21-DEC-1999
 mRNA sequence.

ACCESSION AI760651
 VERSION
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 This 5' resequenced clone has no previous 5' data to verify this
 new read against

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-x@mail.nih.gov
Tissue Procurement: Christopher Mokaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 567 Std Error: 0.00
Seq primer: -40UP from Gibco.
Location/Qualifiers
1. .455
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2398336"
/tissue_type="2 pooled tumors (clear cell type)"
/lab_host="DH108"
/clone_lib="NCI CGAP Kid12"
/notes="Organ: kidney; Vector: pTTT3D-Pac (Pharmacia) with
a modified polylinker; Site:1: Not 1; Site 2: Eco RI;
Plasmid DNA from the normalized library NCI CGAP Kid5 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(clonesID 1323912-1325931, 1471368-1472903 and
1492104-1493255). Subtraction by Bento Soares and M.
Fatima Bonaldo."

RESULT 4	450 bp	mRNA	linear	EST 09-MAR-2000
AW051661/c	450 bp	mRNA	linear	EST 09-MAR-2000
LOCUS	wx27b06.x1	NCI_CGAP_Kid11	Homo sapiens	cDNA clone IMAGE:2544851 3'
DEFINITION	mRNA sequence.			
ACCESSION	AW051661	1	(bases 1 to 450)	
VERSION	AW051661.1	GI:5913892		
KEYWORDS	EST.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	1	(bases 1 to 450)		
TITLE	NCI-CGAP	http://www.ncbi.nlm.nih.gov/ncicgap.		
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),			
COMMENT	Tumor Gene Index			
	Unpublished (1997)			
	Contact: Robert Strausberg, Ph.D.			
	Email: cgapbs-r@mail.nih.gov			
	Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.			
	Emmert-Buck, M.D., Ph.D.			
	cDNA Library Preparation: M. Bento Soares, Ph.D.			
	cDNA Library Arrayed by: Greg Lennon, Ph.D.			
	DNA Sequencing by: Washington University Genome Sequencing Center			
	Clone distribution: NCI-CGAP clone distribution information can be			
	found through the I.M.A.G.E. Consortium/LLNL at:			
	www-bio.llnl.gov/bbrp/image/image.html			
	Insert length: 540	Std Error: 0.00		
	Seq primer: -400P from Gibco.			
FEATURES	Location/Qualifiers			
source	1..450			
	/organism="Homo sapiens"			
	/mol_type="mRNA"			
	/db_xref="taxon:9606"			
	/clone="IMAGE:2544851"			
	/lab_host="DH10B"			
	/clone_lib="NCI_CGAP_Kid11"			
	/note="Organ: kidney; Vector: pTVT3D-Pac (Pharmacia) with			
	a modified polylinker; Site 1: Not I; Site 2: Eco RI;			
	Plasmid DNA from the normalized library NCI_CGAP_Kid3 was			
	prepared, and ss circles were made in vitro. Following HAP			
	purification, this DNA was used as tracer in a subtractive			
	hybridization reaction. The driver was PCR-amplified cDNA			
	from a pool of 5,000 clones made from the same library			
	(cloneIds 1322376-1323911, 1456007-1456775, and			
	1500552-1502855). Subtraction by Bento Soares and M.			
	Fatima Bonaldo."			
ORIGIN				
Query Match	18.1%;	Score 448;	DB 2;	Length 450;
Best Local Similarity	100.0%;	Pred. No. 4.1e-62;		
Matches 448;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;
Qy	676	CGCTCTAGGATCCCCGAAAGGACACGGGCGCGAAAGCGGCTGGGCCAGGATCT	735	
Db	450	CGCTCTAGGATCCCCGAAAGGACACGGGCGCGAAAGCGGCTGGGCCAGGATCT	391	
Qy	736	AGAAAGACTGCTGGCGGAGCTCCCTGCCCCCGGGCCCTGCTCATGTGACTCTGTGGA	795	
Db	390	AGAAAGACTGCTGGCGGAGCTCCCTGCCCCCGGGCCCTGCTCATGTGACTCTGTGGA	331	
Qy	796	GAGCTTCGCTTCCCGCGGAGCCCTTCCTGACGGGCTCCACCTCCAGGCACCCGGGGCTCG	855	
Db	330	GAGCTTCGCTTCCCGCGGAGCCCTTCCTGACGGGCTCCACCTCCAGGCACCCGGGGCTCG	271	
Qy	856	GACACCCACCCCGCGGCGGACACTGCTCCCTGGTGGCCCTTAAACCCGGGGCTGAGTCTCG	915	
Db	270	GACACCCACCCCGCGGCGGACACTGCTCCCTGGTGGTCCCTTAAACCCGGGGCTGAGTCTCG	211	
Qy	916	TTAAGATGGCGAAGTGTCGTCGGTCCGGAACACGGGAAACCCCAATCCGCTGCCGACC	975	
Db	210	TTAAGATGGCGAAGTGTCGTCGGTCCGGAACACGGGAAACCCCAATCCGCTGCCGACC	151	

```

QY 976 TCCTGACCCCGCCCGACGAGACTGGGCTCCGACGCGAGCGGCTGCCG 1035
Db 150 TCTTGACCCCGCCCGACGAGACTGGGCTCCGACGCGAGCGGCTGCCG 91
QY 1036 GGACACCGGTGCTGCGAAACGAGGACCTTTGTAACGCGACGTGTTGCTCTTTTGAA 1095
Db 90 GGACACCGGTGCTGCGAAACGAGGACCTTTGTAACGCGACGTGTTGCTCTTTTGAA 31
QY 1096 AAAACAAGATTAATGTTAACTGTC 1123
Db 30 AAAACAAGATTAATGTTAACTGTC 3

RESULT 5
BQ437878 1240 bp mRNA linear EST 24-MAY-2002
LOCUS AGENCOURT_7890066 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:6176570
5', mRNA sequence.
ACCESSION BQ437878
VERSION BQ437878.1 GI:21176954
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1240)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLNL13553 row: C column: 03
High quality sequence stop: 421.
Location/Qualifiers
1..1240
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6176570"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 71"
/notes="Organ: uterus; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dr.
Average insert size 2.1 kb."

FEATURES
source
Query Match 17.8%; Score 440.6; DB 5; Length 1240;
Best Local Similarity 94.4%; Pred.No. 4.9e-61;
Matches 490; Conservative 0; Mismatches 24; Indels 5; Gaps 3;

QY 626 TGCCAGCGCGGAGAACGCGCTGGGGAAACCTTGGTCCGCTCGCGCGTCGCTTAGGA 685
Db 1 TGCCAGCGCGGAGAACGCGCTGGGGAAACCTTGGTCCGCTCGCGCGTCGCTTAGGA 60
QY 686 TCCCCGAAAGAGGACGCGCGCGGAAAGCGGCCAGGCTGGCCAGGATCTAGAAAGACTG 745
Db 61 TCCCCGAAAGAGGACGCGCGCGGAAAGCGGCCAGGCTGGCCAGGATCTAGAAAGACTG 120
QY 746 CTTGGCGCAGGCTCCTTGCCTCCCGCGGCTGTGTATGACTCGTGGAGAGCTCGCTT 805
Db 121 CTTGGCGCAGGCTCCTTGCCTCCCGCGGCTGTGTATGACTCGTGGAGAGCTCGCTT 180
QY 806 CCGCGCGGACCTTCTCTGAGGGGTCCAGTCCAGGACCGCGCGCTCGACACCCAC 865
Db 181 CCGCGCGGACCTTCTCTGAGGGGTCCAGTCCAGGACCGCGCGCTCGACACCCAC 240

```

```

QY 866 CCCCGCGCGGACCTTGCCCTGGTGCCCTTAACCCCGGCGGTAGCTCGTTAAGATGCC 925
Db 241 CCCCGCGCGGACCTTGCCCTGGTGCCCTTAACCCCGGCGGTAGCTCGTTAAGATGCC 300
QY 926 GAAGTGTCCGGTCCCGAACACGCGAAACCCAAAATCCCGCCTGCCGACCTCTCTGACCCC 985
Db 301 GAAGTGTCCGGTCCCGAACACGCGAAACCCAAAATCCCGCCTGCCGACCTCTCTGACCCC 360
QY 986 CGGCGCCACGCGACGACAGACTGGGCTCCCGACGCGGAGCGGCTGCCGCGGACACC-GG 1044
Db 361 CGGCGCCACGCGACGACAGACTGGGCTCCCGACGCGGAGCGGCTGCCGCGGACACC-GG 420
QY 1045 TCGGTGCGAAACGG--AGGACCTTTGTAAAGCCACGCTG--TTTGCTCTCTTTTGAACAAAC 1100
Db 421 TCGGTGCGAAACGGAGGACCTTTGTAAAGCCACGCTGCTGCTCTTTTGAACAAAC 480
QY 1101 AAGAATAAATGTTAAACTGTCTGAAAAGCTTGCGGCC 1139
Db 481 CAGAATAAATGCTCACTGGCTGAAAAGGTCGTGCC 519

RESULT 6
AI685715/c 435 bp mRNA linear EST 07-MAR-2000
LOCUS tu37c07.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2253228 3',
DEFINITION mRNA sequence.
ACCESSION AI685715
VERSION AI685715.1 GI:4897009
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 435)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 562 Std Error: 0.00
Seq primer: -40UP from Gibco.
Location/Qualifiers
1..435
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2253228"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="NCI CGAP Pr28"
/notes="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
with a modified polylinker; Plasmid DNA from the
normalized library NCI_CGAP_Pr22 was prepared, and as
circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (clonesIDs
985608-986759, 1101192-1101959, and 1217928-1220615).
Subtraction by Bento Soares and M. Fatima Bonaldo."

FEATURES
source
Query Match 17.6%; Score 435; DB 1; Length 435;
ORIGIN

```

```

Best Local Similarity 100.0%; Pred. No. 5.1e-60;
Matches 435; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 689 CCGAAAGAGACACGGCGCGAAGCGCGCAGGCTGGCCAGGATCTAGAAAGACTGCCT 748
Db 435 CCGAAAGAGACACGGCGCGAAGCGCGCAGGCTGGCCAGGATCTAGAAAGACTGCCT 376

QY 749 GCGCAGGCTCCTGCGCCCGCGGGCTGCTGTATGAGACTCGTGAGAGCTCGCTTCCC 808
Db 375 GCGCAGGCTCCTGCGCCCGCGGGCTGCTGTATGAGACTCGTGAGAGCTCGCTTCCC 316

QY 809 GCGCAGGCTCCTGCGCCCGCGGGTCCAGCTCCAGGACCGCGGCTCGGACACCCACCCC 868
Db 315 GCGCAGGCTCCTGCGCCCGGGTCCAGCTCCAGGACCGCGGCTCGGACACCCACCCC 256

QY 869 CCGCCCGGACCTGCGCCCTGGTCCCTTAACCCGGCGGTAGCTGTTAAGATGCGAA 928
Db 255 CCGCCCGGACCTGCGCCCTGGTCCCTTAACCCGGCGGTAGCTGTTAAGATGCGAA 196

QY 929 GTGTCCGCTCCGGAACACCGAAACCCCAATCCCGCTGCCGACTCTGACCCCGG 988
Db 195 GTGTCCGCTCCGGAACACCGGAACCCCAATCCCGCTGCCGACTCTGACCCCGG 136

QY 989 CCCACCGGACGACAGACTGGGCTCCCGACGCGAGCGGCTGCGCGGACACCGGTGCG 1048
Db 135 CCCACCGGACGACAGACTGGGCTCCCGACGCGAGCGGCTGCGCGGACACCGGTGCG 76

QY 1049 TCGGAACCGAGGACCTTTGTAAACGCGCGTGTCTCTTTTGAAGAAACGAATAA 1108
Db 75 TCGGAACCGAGGACCTTTGTAAACGCGCGTGTCTCTTTTGAAGAAACGAATAA 16

QY 1109 ATGTGTTAAACTGTC 1123
Db 15 ATGTGTTAAACTGTC 1

RESULT 7
A1733521/c
LOCUS
DEFINITION
Ov05609.x5 NCI_CGAP_Kid3 Homo sapiens cDNA clone IMAGE:1636456 3',
mRNA sequence.
A1733521
VERSION
A1733521.1 GI:5054682
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCI/NIDR-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute / National Institute of Dental Research,
Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-rc@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.lnl.gov/bbrp/image/image.html

This read is a RESSEQUENCE of a previously sequenced human clone
Original clone citation: National Cancer Institute, Cancer Genome
Anatomy Project (CGAP), Tumor Gene Index
This read has been verified (found to hit its original self in the
correct orientation)
Seq primer: -40UP from Gibco.
Location/Qualifiers
1. .435
/organism="Homo sapiens"

FEATURES
source

```

```

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1636456"
/lab_host="DH10B"
/clone_lib="NCI CGAP Kid3"
/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer,
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT7T3 vector. mRNA
source: 2 pooled kidneys. Library went through one round
of normalization. Library constructed by Bento Soares and
M. Fatima Bonaldo. "

ORIGIN

Query Match
Best Local Similarity 100.0%; Pred. No. 5.1e-60;
Matches 435; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 689 CCGAAAGAGACACGGCGCGAAGCGCGCAGGCTGGCCAGGATCTAGAAAGACTGCCT 748
Db 435 CCGAAAGAGACACGGCGCGAAGCGCGCAGGCTGGCCAGGATCTAGAAAGACTGCCT 376

QY 749 GCGCAGGCTCCTGCGCCCGCGGGCTGCTGTATGAGACTCGTGAGAGCTCGCTTCCC 808
Db 375 GCGCAGGCTCCTGCGCCCGCGGGCTGCTGTATGAGACTCGTGAGAGCTCGCTTCCC 316

QY 809 GCGCAGGCTCCTGCGCCCGGGTCCAGCTCCAGGACCGCGGCTCGGACACCCACCCC 868
Db 315 GCGCAGGCTCCTGCGCCCGGGTCCAGCTCCAGGACCGCGGCTCGGACACCCACCCC 256

QY 869 CCGCCCGGACCTGCGCCCTGGTCCCTTAACCCGGCGGTAGCTGTTAAGATGCGAA 928
Db 255 CCGCCCGGACCTGCGCCCTGGTCCCTTAACCCGGCGGTAGCTGTTAAGATGCGAA 196

QY 929 GTGTCCGCTCCGGAACACCGAAACCCCAATCCCGCTGCCGACTCTGACCCCGG 988
Db 195 GTGTCCGCTCCGGAACACCGAAACCCCAATCCCGCTGCCGACTCTGACCCCGG 136

QY 989 CCCACCGGACGACAGACTGGGCTCCCGACGCGAGCGGCTGCGCGGACACCGGTGCG 1048
Db 135 CCCACCGGACGACAGACTGGGCTCCCGACGCGAGCGGCTGCGCGGACACCGGTGCG 76

QY 1049 TCGGAACCGAGGACCTTTGTAAACGCGCGTGTCTCTTTTGAAGAAACGAATAA 1108
Db 75 TCGGAACCGAGGACCTTTGTAAACGCGCGTGTCTCTTTTGAAGAAACGAATAA 16

QY 1109 ATGTGTTAAACTGTC 1123
Db 15 ATGTGTTAAACTGTC 1

RESULT 8
BF446188/c
LOCUS
DEFINITION
BF446188
mRNA sequence.
BF446188
VERSION
BF446188.1 GI:11511326
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-rc@mail.nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.

```


[illegible]

```
715 GGCAGGCTGGCCAGGATCTAGAAAGACTGCTGGCGCAGGCTCCCTGCCCGCGGGC 774
Db |||||
415 GGCAGGCTGGCCAGGATCTAGAAAGACTGCTGGCGCAGGCTCCCTGCCCGCGGGC 356
775 CTGCTGTCTATGGACTCGTGAGAGCTCGCTTTCCCGCGGACACCTTCTCTGCAGGGTCCA 834
Db |||||
355 CTGCTGTCTATGGACTCGTGAGAGCTCGCTTTCCCGCGGACACCTTCTCTGCAGGGTCCA 296
835 CGTCCAGGACACCGCGGCTCGGACACCCACACCCCGCGGCGGACCTGCGCCTGGGTGCC 894
Db |||||
295 CGTCCAGGACACCGCGGCTCGGACACCCACACCCCGCGGCGGACCTGCGCCTGGGTGCC 236
895 CTTAAACCCGGGCTAGCTCGTTAAGATGGCGAAGTGTCCGTCGCGAACACCGGAAAC 954
Db |||||
235 CTTAAACCCGGGCTAGCTCGTTAAGATGGCGAAGTGTCCGTCGCGAACACCGGAAAC 176
955 CCAAAATCCCGCCTCGCCGACCTCTGACACCCCGCGGCGGACGACAGACTGGGCTTC 1014
Db |||||
175 CCAAAATCCCGCCTCGCCGACCTCTGACACCCCGCGGCGGACGACAGACTGGGCTTC 116
1015 CGACCGGACGCGCTGCGCGGACACCGGTGCGTGCAGAAACCGAGGACCTTTGTAAAGC 1074
Db |||||
115 CGACCGGACGCGCTGCGCGGACACCGGTGCGTGCAGAAACCGAGGACCTTTGTAAAGC 56
1075 CAGCTGTTCCTCTTTTGAAGAAACAGAGTAATGTGTAAACTGTCTGAAA 1129
Db |||||
55 CAGCTGTTCCTCTTTTGAAGAAACAGAGTAATGTGTAAACTGTCTGAAA 1

RESULT 12
BE048839/c
LOCUS
DEFINITION hr54e04.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:3132318 3',
mRNA sequence.
ACCESSION BE048839
VERSION BE048839.1 GI:8365894
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 404)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -40UP from Gibco.
Location/Qualifiers
1. .404
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3132318"
/lab_host="DH10B"
/clone_lib="NCI CGAP Kid11"
/notes="Organ: kidney; Vector: p7T3D-Pac (Pharmacia) with
a modified polylinker; Site:1: Not 1; Site 2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP_Kid3
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library"

FEATURES
source

722 CTGGGCGAGGATCTAGAAAGACTGCTGGCGCAGGCTCCCTGCCCGCGGGCTGCTGT 781
Db |||||
404 CTGGGCGAGGATCTAGAAAGACTGCTGGCGCAGGCTCCCTGCCCGCGGGCTGCTGT 345
782 CATGAGCTCGTGGAGAGCTCGCTTCCCGCGGACCCCTTCTCTGCAGGGTCCAG 841
Db |||||
344 CATGAGCTCGTGGAGAGCTCGCTTCCCGCGGACCCCTTCTCTGCAGGGTCCAG 285
842 GCACCGCGGCTCGGACACCCACACCCCGCGGCGGACCTGCGCTGGGTGCCCTTAAAC 901
Db |||||
284 GCACCGCGGCTCGGACACCCACACCCCGCGGCGGACCTGCGCTGGGTGCCCTTAAAC 225
902 CGGCGGTAGCTCGTTAAGATGGCGAAGTGTCCGTCGGAACACCGGAAACCCCAAATC 961
Db |||||
224 CGGCGGTAGCTCGTTAAGATGGCGAAGTGTCCGTCGGAACACCGGAAACCCCAAATC 165
962 CGGCTGCGCGACCTCTGACACCCCGCGGCGGACGACAGCTGGGCTCCCGACGC 1021
Db |||||
164 CGGCTGCGCGACCTCTGACACCCCGCGGCGGACGACAGCTGGGCTCCCGACGC 105
1022 GCAGCGCGCTGCGGCGACACCGGTGCGTGCAGAAACCGAGGACCTTTGTAAACGCGTGT 1081
Db |||||
104 GCAGCGCGCTGCGGCGACACCGGTGCGTGCAGAAACCGAGGACCTTTGTAAACGCGTGT 45
1082 TTGCTCTTTTGAAGAAACAGAGTAATGTGTAAACTGTCTG 1125
Db |||||
44 TTGCTCTTTTGAAGAAACAGAGTAATGTGTAAACTGTCTG 1

RESULT 13
AI498080/c
LOCUS
DEFINITION tm66a04.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2163054 3',
mRNA sequence.
ACCESSION AI498080
VERSION AI498080.1 GI:4390062
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 389)
NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/BTGP), Tumor Gene Index
Unpublished (1998)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Insert Length: 516 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 378.
Location/Qualifiers
1. .389
/organism="Homo sapiens"

FEATURES
source
```

```
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2163054"
/tissue_type="anaplastic oligodendroglioma"
/lab_host="DH10B"
/lab_lib="NCI CGAP Brn25"
/notes="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCATCTGAAGTGGAGCGCGCATAGTCTTTTCTTTTCTTTTCTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
Library is normalized, and was constructed by Bento
Soares and M.Fatima Bonaldo."
```

ORIGIN

```
Query Match 14.7%; Score 362.8; DB 1; Length 389;
Best Local Similarity 99.2%; Pred. No. 2e-48;
Matches 375; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 752 GCAGGCTCCCTGCCCCCGGGGCTGCTCATGGACTCGTGAGAGCTCGCTCCCGG 811
DB 389 GCAGGCTCCCTGCCCCCGGGGCTGCTCATGGACTCGTGAGAGCTCGCTCCCGG 330
QY 812 CGGACCTTCTCGAGGGTCCACGTCACGACCGCGCTCGGACACCCACCCCGG 871
DB 329 CGGACCTTCTCGAGGGTCCACGTCACGACCGCGCTCGGACACCCACCCCGG 271
QY 872 CCGGGACCTCGCTCGCTGCCCCCTTAACCCGGCGGTAGCTCGTTAAGATGCGGAAGTG 931
DB 270 CCGGGACCTCGCTCGCTGCCCCCTTAACCCGGCGGTAGCTCGTTAAGATGCGGAAGTG 211
QY 932 TCCGTTCCGGAACACCGGAAACCCCAATCCCGCTGCCGACCTCTGACCCCGGCC 991
DB 210 TCCGTTCCGGAACACCGGAAACCCCAATCCCGCTGCCGACCTCTGACCCCGGCC 151
QY 992 CACGGACACAGACTGGGCTCCCGACGCGAGCGGCTGCGGACACCGGTGGGTGC 1051
DB 150 CACGGACACAGACTGGGCTCCCGACGCGAGCGGCTGCGGACACCGGTGGGTGC 91
QY 1052 GAAACGGAGGACCTTTGTAAACGACAGTGTGTCTTTTGTAAACAAAGAAATG 1111
DB 90 GAAACGGAGGACCTTTGTAAACGACAGTGTGTCTTTTGTAAACAAAGAAATG 31
QY 1112 TGTAAACTGTCTGAAA 1129
DB 30 TGTAAACTGTCTGAAA 13
```

RESULT 14

```
BC039100 3491 bp mRNA linear HTC 04-MAR-2003
LOCUS Homo sapiens, clone IMAGE:4823290, mRNA.
DEFINITION BC039100
ACCESSION BC039100
VERSION BC039100.1 GI:25058236
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 3491)
Strausberg, R.
Direct Submission
Submitted (01-NOV-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: gcgaps@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
```

Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulesged, H.,
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAC Plate: 71 Row: h Column: 3
This clone has the following problem: retained intron.

FEATURES

Source

```
1..3491
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4823290"
/tissue_type="Testis"
/clone_lib="NIH MGC_97"
/lab_host="DH10B"
/note="Vector: pBluescript"
```

ORIGIN

```
Query Match 13.4%; Score 332; DB 3; Length 3491;
Best Local Similarity 71.1%; Pred. No. 1.1e-43;
Matches 495; Conservative 0; Mismatches 195; Indels 6; Gaps 4;

QY 1508 AAAAAATAGGGGTACTAGCCAGCGGGTGGCTGGCCCTGTATATCCAGACCTTGGGAG 1567
DB 832 AAAAAAGGATAGGGGCTGGGCACTGCTGATGATATCCAGACCTTGGGAG 891
QY 1568 GCGAGGACGCGGATCCTTGGGTGAGGTTCAGAGTTCAGACCGCTGTCACATGGGTG 1627
DB 892 GCGAGGACGCGGATCCTTGGGTGAGGTTCAGAGTTCAGACCGCTGTCACATGGGTG 950
QY 1628 AAACGCGGCTCTACCAAAAATAATAAACTTAGCTGAGTGTGTCAGCGCATGCTGTAA 1687
DB 951 AAACCCGCTCTCTACCAAAAATAATAAACTTAGCTGAGGATGTGTCAGCGCATGCTGTAA 1009
QY 1688 TCCAGCAACTCAGGAGGCTGAGGCA - GAGAAATCGCTTGAACCTGGAGCGGAGGTGC 1746
DB 1010 TCCAGCTACTAAGAGCCGAGGAGGAAATCACTTGAACCTGGAGCGGAGGATGC 1069
QY 1747 AGTGAGCTGAGATCGTACACCTCCAGCTTGGCGGACAGCGAGACTTGCCTTA 1806
DB 1070 AGTGAGCTGAGATCGTCCGCTGCTCCTCCAGCATGGGCAACAGAGCGAGACTTGTCTCA 1129
QY 1807 AAATAAATAAATAAATTTTAAAAAAA - ATAGGGGTACTAATATCTACCTTAAAGGAT 1863
DB 1130 AAAAGGAAAAAAGAAAGAAATAGGATTTTTTTTTTTTGTACTGTGTACTCTTT 1189
QY 1864 GAGGTTAAATTAAGTACACACATAAGCCCTAGCGAGTGGCTTATGCTGTATCTCAA 1923
DB 1190 GATGTTAAATGAAAAAATAAATAAGCCAGGACATTTGGCTCAGCCTGTATCCAG 1249
QY 1924 CACTTTGGGAGTCTGTGGGAGGAGTACCTTGAAGCCAGGAGTTTGAGACTAGTCTGGG 1983
DB 1250 CACTTTGGGAGTCTGTGGGAGGAGTACCTTGAAGCCAGGAGTTTGAGACTAGTCTGGG 1309
QY 1984 CAACAGACATCTCTATAGTTGTTGTTTGTGTTTACCAGTGTGGTGGTGTG 2043
DB 1310 CAACATGATGAAACCCCGCTTTTACTTAAATAATACAAATAGCTGGGCAATGTTGGTATG 1369
QY 2044 CACTTCAGTCCAGCTTACTAGGAGGCTGAGGTGGGAGGACTGCTTGCAGCCAGGAGGT 2103
DB 1370 CACTGTATCTCAGCTACTTGGAGGCTGAGGAGGAGAAATGCTTGAACCCCGGAGGT 1429
QY 2104 CGAGGCTGAGTGAAGCCATGTTGTGCTCCTGACCTGCTCCAGCTGGGCAACACAGCAAGAC 2163
```

Db 1430 GGAGGTTGAGGCGGAGAGTTGGCCACTGCTCCTGGGTGACAGAGTGAGAC 1489

Qy 2164 CTGTCTCAAAAACAAAAGCATCTATAAA 2199

Db 1490 CCGTCTCAAAATATAATAATAATAATAATA 1525

RESULT 15

CD557847

LOCUS CD557847 897 bp mRNA linear EST 11-JUN-2003

DEFINITION AGENCOURT_14423258 NIH_MGC_180 Homo sapiens cDNA clone IMAGE:30390116 5', mRNA sequence.

ACCESSION CD557847

VERSION CD557847.1 GI:31583915

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 897)

REFERENCE NIH-MGC <http://mgs.nci.nih.gov/>.

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE Unpublished (1999)

JOURNAL

COMMENT Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgabs-remail.nih.gov
cDNA Library Preparation: Dr. Michael Brownstein
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: NDAM464 row: k column: 21
High quality sequence stop: 636.
Location/Qualifiers
1..897
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30390116"
/lab_host="DH10B-Ton A (T1 and T5 phage resistances)"
/notes="Organ: Testis; Vector: pCMV-SPORT6.1; Site 1: NotI; Site 2: EcoRV (destroyed); Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.68 kb. Library was constructed by (Invitrogen). Note: this is a NIH_MGC Library."

FEATURES
source

ORIGIN

Query Match 12.4%; Score 305.2; DB 6; Length 897;
Best Local Similarity 68.9%; Pred. No. 3e-39;
Matches 495; Conservative 0; Mismatches 208; Indels 15; Gaps 5;

Qy 1472 TTGTTTAACCTCTTTGACCTCAGCTTCTTCATTACAAAATAGGGGTACTAGCCAGCGG 1531

Db 34 TTGTTTTCCTCACTACATTGATGTTGGGAATTACAGATAACGGAAGCCGGCTGGG 93

Qy 1532 GGGTGGCTCGCGCTGTAATCCAGACATTGGGAGCGCCAGGACCGGATCACTTGAG 1591

Db 94 TGGTGGCTCAGCGCTGTAAACCCCAACACTTTGGGAGGCCAAGCGCGGATCACCTGAG 153

Qy 1592 GTCAGAGTTTCAGACAGCTGGTCAACATGGGTGAACCGCGGCTCTACCAAAAATAT 1651

Db 154 GTCAGAGTTTCAGAGATTAGTCTGGCCAACT-GGTGAACCCCATCTCTACTAAAAATAC 212

Qy 1652 AAAAACTAGCTGAGTGTGGTAGCGCATGCTGTAATCCAGCAACTCAGGAGGCTGAGG 1711

Db 213 GAAA--TTAGCCAGGTGTGGTGGCACACATCTGTAGTCCCGACTCTCTGGAGGCTGAGA 270

Qy 1712 CA-GAGATCGCTTGAACTCGGAGCGGAGGTTGCAGTGAGCTGAGATCGTACCACTGC 1770

Db 271 CAGGAGATCGCTTGAACTCGGAGGAGTGGAGGTTGCAGTGAGCTGAGATCACCACCTGC 330

Qy 1771 ACTCCAGCTTTGGCGACAGAGCGAGACTCTGCTCTTAAATAATAATAATAATTTTAAAA 1830

Db 331 ACTCCAGCTTTGGCGAGAGAGAGTGTCTCCATCTCAAAAAA-----AAAA 381

Qy 1831 AAAATAGGGTACTAATATCTACCTTTAAGGATGAGGTTAAATTAAGTACACACATAAG 1890

Db 382 GATAGAAGCAATAAGCATGGTGCAATCAAAATTCTGGCAAGCATTAATAATCAGGATGCA 441

Qy 1891 CCTAGCGCAGTGGCTTATGCTGTAATCTCAACACTTTGGGAGTCTGTGGCGGAGGAT 1950

Db 442 GCTGGGACGGTGGCTCAGCCCTGTAAATCCAGCACTTTGGGAGGCAAGGTGGGAGAT 501

Qy 1951 CACTTGAGCCAGGAGTTTGAGACTAGTCTGGGCAACAGAGACATGTCTCTATAGT--TG 2008

Db 502 CACTTGAGGTCAAGAAATTTGAGAGGAGCCTGGCCAGCATGGCAAAACCCCATCTGTACTT 561

Qy 2009 TGTGTTGGTTTGTGTTTACAGGTGTGGTGTGTCACCTGCAGTCCCGAGCTACTAGGGA 2068

Db 562 AAAATACAAAAAATTTAGCTGGGCGTGTGTGTCACACCTGTAATCCAGCTACTTTGGGA 621

Qy 2069 GGCTGAGGTGGGAGGACTGCTGAGCCAGGAGGTGAGGCTGAGTGGAGTGGAGTGGATGTT 2128

Db 622 GGCTGAGGTGGGAGAAATTGCTTGAACCTGGAGGTGGAGGTTGCAGTGAGCTGAGATCTT 681

Qy 2129 GCCACTGCACTCCAGCTGGGCAACACAGAGACCTTGTCTCAAAAAACAACAAAA 2186

Db 682 GCCACTGCACTCCAGGCTGGGCAACAGAGTGAAGCCATGTCTCAAAAAATAAAAAATA 739

Search completed: August 12, 2005, 04:53:20
Job time : 7901 secs

Qy 61 CATTTCAGTAAAGAAAAACGAGCAACGGAAACAGTTCGCGAGCGCGCAGACCCCGCA 120

Db 61 CATTTCAGTAAAGAAAAACGAGCAACGGAAACAGTTCGCGAGCGCGCAGACCCCGCA 120

```

US-09-891-711-2
; Sequence 2, Application US/09891711
; Patent No. US20020982404A1
; GENERAL INFORMATION:
; APPLICANT: Pharmacia AB
; TITLE OF INVENTION: Promoter Sequences
; FILE REFERENCE: 00130
; CURRENT APPLICATION NUMBER: US/09/891,711
; CURRENT FILING DATE: 2001-06-26
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2469

```

121 GGGCCGGGTACAGGACGCTGTGTCACAAACAGCGCGAGGCCCGCGGCCACCTCC 180
121 GGGCCGGGTACAGGACGCTGTGTCACAAACAGCGCGAGGCCCGCGGCCACCTCC 180
181 CCAGCCGGCCCGGCCCGGCCCGCTCGCTCGGGGCTTCGAGCAACCGGACACCT 240
181 CCAGCCGGCCCGGCCCGGCCCGCTCGCTCGGGGCTTCGAGCAACCGGACACCT 240
241 GAGCGAGCGGCCCGCCACCTAGGCGGAGCGGCTCGGGAGGCCCGCGCGGGCTG 300
241 GAGCGAGCGGCCCGCCACCTAGGCGGAGCGGCTCGGGAGGCCCGCGCGGGCTG 300
301 ACCTACTCGCGCGCGGAGCTCAGGGCGGCGGCGCGGGATGACGGCGCTCCGCG 360
301 ACCTACTCGCGCGCGGAGCTCAGGGCGGCGGCGCGGGATGACGGCGCTCCGCG 360
361 CGAACACGCTCGGGCACTCCATTCGGGGCTGTTTACTCCCACTCTCGCGAGACTGGCG 420
361 CGAACACGCTCGGGCACTCCATTCGGGGCTGTTTACTCCCACTCTCGCGAGACTGGCG 420
421 GCCGGCGAGCGAGGCCCAAGCTGGGAGCTCAGCTCGGCCGACCGAGGTGCCCTGTC 480
421 GCCGGCGAGCGAGGCCCAAGCTGGGAGCTCAGCTCGGCCGACCGAGGTGCCCTGTC 480
481 TGTCCCGCGCTCCCGGGCTTGGTGGCGCTCTGGACGCGGTGGGCGGACCGACCG 540
481 TGTCCCGCGCTCCCGGGCTTGGTGGCGCTCTGGACGCGGTGGGCGGACCGACCG 540
541 CGGAGGATGGAAGAGTGTCTCGACATTTTCGGCGCGGGCGGGTGGCAGGGTGG 600
541 CGGAGGATGGAAGAGTGTCTCGACATTTTCGGCGCGGGCGGGTGGCAGGGTGG 600
601 AAGCGAGGGCGTGGCGAGAGTGTCTCGAGCGGGGAGAACCGCTGGGGAAACCTTG 660
601 AAGCGAGGGCGTGGCGAGAGTGTCTCGAGCGGGGAGAACCGCTGGGGAAACCTTG 660
661 GTCGGCTCTGCGGTGCTCTAGGATCCCGGAAAGAGCACGGCGCGGAGGCCAG 720
661 GTCGGCTCTGCGGTGCTCTAGGATCCCGGAAAGAGCACGGCGCGGAGGCCAG 720
721 GCTGGGCGAGGATCTAGAAAGTGTCTGGCGAGGCTCCCTGCCCGCGGGCTGCTG 780
721 GCTGGGCGAGGATCTAGAAAGTGTCTGGCGAGGCTCCCTGCCCGCGGGCTGCTG 780
781 TCATGAGCTGTGGAGAGTGTCTTCCGGCGGGAACCTTCTCTGAGGGTTCACGTCCA 840
781 TCATGAGCTGTGGAGAGTGTCTTCCGGCGGGAACCTTCTCTGAGGGTTCACGTCCA 840
841 GGCACCGGGCTCGGACACCCACCCCGGGCGGACCTGCTGGGTGCCCTTAAAC 900
841 GGCACCGGGCTCGGACACCCACCCCGGGCGGACCTGCTGGGTGCCCTTAAAC 900
901 CCGGGCGGTAGCTCGTTAAGTGGCAAGTGTCCGGTTCGGGAACACCGCGAAACCCCAAT 960
901 CCGGGCGGTAGCTCGTTAAGTGGCAAGTGTCCGGTTCGGGAACACCGCGAAACCCCAAT 960
961 CCGGCTGCCGACCTCTGACCCCGGGCCCAAGGAGCTGCTGGGCTCCGACG 1020
961 CCGGCTGCCGACCTCTGACCCCGGGCCCAAGGAGCTGCTGGGCTCCGACG 1020
1021 CGCAGCGCTGCGGGACACCGGTGCTGCGAAACCGAGGACCTTTGTAACGCGACGTG 1080
1021 CGCAGCGCTGCGGGACACCGGTGCTGCGAAACCGAGGACCTTTGTAACGCGACGTG 1080
1081 TTTGCTCTTTTGAAGAAACAGAAATGTTTAACTGTCTGAAAGCTTGGCGCT 1140
1081 TTTGCTCTTTTGAAGAAACAGAAATGTTTAACTGTCTGAAAGCTTGGCGCT 1140
1141 AAAAGATGTCTGGGTGATGTAGATGATGATGATGATGATGATGATGATGATGATGAT 1200
1141 AAAAGATGTCTGGGTGATGTAGATGATGATGATGATGATGATGATGATGATGATGAT 1200
1201 CCCGGACTCCGACCTAGCAGGGGACTGAAAGCGTCTTCAGGTACTGCTGGTGGGC 1260

1201 CCGGACTCCGTACGGCACTAGCAGGGGACTGAAGCGTCTTCAGGTACTGCTGGTGGGC 1260
1261 GGTGATGCGCTACAGGCCGATCAGACAGTTTTGTGTCTTCTGGAACTTGACACTGCACCA 1320
1261 GGTGATGCGCTACAGGCCGATCAGACAGTTTTGTGTCTTCTGGAACTTGACACTGCACCA 1320
1321 CGGTAACTGCTGAATGCAACCAATATACAGATCAGCGCATCATCTTCTTCAACATGA 1380
1321 CGGTAACTGCTGAATGCAACCAATATACAGATCAGCGCATCATCTTCTTCAACATGA 1380
1381 TTTTAAACACAGTTTGAATATATGTTGGAATAATGAGATCACAATTAACCATACCCAC 1440
1381 TTTTAAACACAGTTTGAATATATGTTGGAATAATGAGATCACAATTAACCATACCCAC 1440
1441 CTCAGGCTCTACTCTCGTAATTTTGGAGCGGTGTTTAACTCTTTTGTACCTCAGCTTCT 1500
1441 CTCAGGCTCTACTCTCGTAATTTTGGAGCGGTGTTTAACTCTTTTGTACCTCAGCTTCT 1500
1501 TCATTTACAAAATAGGGTACTAGCCAGCGGGGTGGCTCGCGCTGTATCCAGCACT 1560
1501 TCATTTACAAAATAGGGTACTAGCCAGCGGGGTGGCTCGCGCTGTATCCAGCACT 1560
1561 TGGGAGGCGGAGGCGAGCGGATCACTTGAAGTTCAGAACTTTCAGACCAAGCTTGTCAAC 1620
1561 TGGGAGGCGGAGGCGAGCGGATCACTTGAAGTTCAGAACTTTCAGACCAAGCTTGTCAAC 1620
1621 ATGGGTGAAACCGCGCTCTACCAAAAATATAAAAATTTAGCTGAGTGTGAGCGCATG 1680
1621 ATGGGTGAAACCGCGCTCTACCAAAAATATAAAAATTTAGCTGAGTGTGAGCGCATG 1680
1681 ACTGTAATCCAGCAACTCAGGAGCTGAGGAGAGAGATCGCTTGAACCTTGGAGGCGGA 1740
1681 ACTGTAATCCAGCAACTCAGGAGCTGAGGAGAGAGATCGCTTGAACCTTGGAGGCGGA 1740
1741 GGTTCAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1800
1741 GGTTCAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1800
1801 GCTTAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 1860
1801 GCTTAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 1860
1861 GATGAGGTAAATTAAGTACACATAGCCCTAGCGAGTGGCTTATGCCGTGAATCT 1920
1861 GATGAGGTAAATTAAGTACACATAGCCCTAGCGAGTGGCTTATGCCGTGAATCT 1920
1921 CAACACTTTGGGAGTCTGTGGGCGGAGGATCACTTGAAGTGGCTTATGCCGTGAATCT 1980
1921 CAACACTTTGGGAGTCTGTGGGCGGAGGATCACTTGAAGTGGCTTATGCCGTGAATCT 1980
1981 GGGCAACAGAGACATGCTCTATAGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 2040
1981 GGGCAACAGAGACATGCTCTATAGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 2040
2041 GTGCACTTCGAGTCCAGCTACTAGGAGGCTGAGGTGGAGGACTGCTTGAAGTGGCTG 2100
2041 GTGCACTTCGAGTCCAGCTACTAGGAGGCTGAGGTGGAGGACTGCTTGAAGTGGCTG 2100
2101 GGTTCAGGCTGAGTGGAGGATGATGTTGCCACTGCACTCCAGCTTGGGCAACACAGCAA 2160
2101 GGTTCAGGCTGAGTGGAGGATGATGTTGCCACTGCACTCCAGCTTGGGCAACACAGCAA 2160
2161 GACCTTGTCTCAAAAACAAAACAAAACAAAACAAAACAAAACAAAACAAAACAAAAC 2220
2161 GACCTTGTCTCAAAAACAAAACAAAACAAAACAAAACAAAACAAAACAAAACAAAAC 2220
2221 CAATATGCTGGTGAATCTTGAATCTTCTGACTCAGATCTCATAGATTTTCTGA 2280
2221 CAATATGCTGGTGAATCTTGAATCTTCTGACTCAGATCTCATAGATTTTCTGA 2280
2281 ACTTTTGGAGAACTCCTTGGCTCTCTGCAATTTGCAACCGGTCAAGGCACTCCCTTCTGCC 2340

Db 2281 ACTTTGGAGAACTCTGCTCTCTGCAATTTGCAAAACCGTCFAAGGCACTCCCTCTGCCC 2340
QY 2341 ACCACAAAGCAATTTGATTTTAACTTGACTATGTCTCTCTCTTCCAACTTTAGGTAAA 2400
Db 2341 ACCACAAAGCAATTTGATTTTAACTTGACTATGTCTCTCTTCCAACTTTAGGTAAA 2400
QY 2401 TTAATCTTGGTCAGGGTTCTCTGAAACAGCCCTTTTACTGCACTATGTCATTTGAATFACATGCG 2460
Db 2401 TTAATCTTGGTCAGGGTTCTCTGAAACAGCCCTTTTACTGCACTATGTCATTTGAATFACATGCG 2460
QY 2461 CCTACAGCT 2469
Db 2461 CCTACAGCT 2469

RESULT 2
US-10-829-118-2
; Sequence 2, Application US/10829118
; Publication No. US20040191759A1
; GENERAL INFORMATION:
; APPLICANT: Pharmacia AB
; TITLE OF INVENTION: Promoter Sequences
; FILE REFERENCE: 00130
; CURRENT APPLICATION NUMBER: US/10/829,118
; CURRENT FILING DATE: 2004-04-20
; PRIOR APPLICATION NUMBER: US/05/891,711
; PRIOR FILING DATE: 2001-06-26
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2469
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-829-118-2

Query Match 100.0%; Score 2469; DB 19; Length 2469;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2469; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGAGTCTGTCTGGAGGCTCCGGCCAGAGCAGGGCGTATTGTTTCACTCGGTGAATGCT 60
Db 1 TTGAGTCTGTCTGGAGGCTCCGGCCAGAGCAGGGCGTATTGTTTCACTCGGTGAATGCT 60

QY 61 CATTTACGTTAAGAAACACAGCAACGGAACGCTGCCGAGCGCGAGACCCCGCA 120
Db 61 CATTTACGTTAAGAAACACAGCAACGGAACGCTGCCGAGCGCGAGACCCCGCA 120

QY 121 GGGCCGGGTACAGGCAACGCTGTGTCCAAACAAAGCGCGGAGCGCCCGCCACCTCCC 180
Db 121 GGGCCGGGTACAGGCAACGCTGTGTCCAAACAAAGCGCGGAGCGCCCGCCACCTCCC 180

QY 181 CCGACCCGGCCCGCCCGCCCGCTCGCTCGGGGCTTCGAGCAACCGGCAACCT 240
Db 181 CCGACCCGGCCCGCCCGCCCGCTCGCTCGGGGCTTCGAGCAACCGGCAACCT 240

QY 241 GAGCGAGCGGCGCCCAACCGCTAGGCGAGCGGGTCGGGAGCGCGCGCGCGGCTG 300
Db 241 GAGCGAGCGGCGCCCAACCGCTAGGCGAGCGGGTCGGGAGCGCGCGCGCGGCTG 300

QY 301 ACCTACTGCGCGCGCGGAGCTCAGGCGCGCGCGCGCGCGGATGACGGCGCTCCGCGG 360
Db 301 ACCTACTGCGCGCGCGGAGCTCAGGCGCGCGCGCGCGCGGATGACGGCGCTCCGCGG 360

QY 361 CGAACAGCTTGGGCACTCCATTTGGGGCTGTTTACTTCCCAACTCTCGCGAGACTGGGCG 420
Db 361 CGAACAGCTTGGGCACTCCATTTGGGGCTGTTTACTTCCCAACTCTCGCGAGACTGGGCG 420

QY 421 GCGGGCGCAGGAGGCGCCACAGCTGGAGCGCTCAGCTCCGCGGACCCAGCGTGCCTGTC 480
Db 421 GCGGGCGCAGGAGGCGCCACAGCTGGAGCGCTCAGCTCCGCGGACCCAGCGTGCCTGTC 480

QY 481 TGTCCCGCTCCCGGGCTTGGTGGCGCTCTGGAGCGCGCTGGGCGAGCGGACCAACGC 540
Db 481 TGTCCCGCTCCCGGGCTTGGTGGCGCTCTGGAGCGCGCTGGGCGAGCGGACCAACGC 540

Db 481 TGTCCCGCTCCCGGGCTTGGTGGCGCTCTGGAGCGCGCTGGGCAACCGGACCAACGC 540
QY 541 CCGGAGGATGAGACGAAGTGTCTCGGACATTTTCGGCGCGCGGGCGCGGTGGCAGGCTGG 600
Db 541 CCGGAGGATGAGACGAAGTGTCTCGGACATTTTCGGCGCGCGGGCGCGGTGGCAGGCTGG 600
QY 601 AAGCGGAGGGGCTGGCCAGAGCTGTCAGGCGCGGAGAAACGCGTCTGGGGGAAACCTTTG 660
Db 601 AAGCGGAGGGGCTGGCCAGAGCTGTCAGGCGCGGAGAAACGCGTCTGGGGGAAACCTTTG 660
QY 661 GTCCGCTCTGCGGCTCGCTTAGGATCCCGGAAAGGAGACGCGGCGGAAAGCGGCCAG 720
Db 661 GTCCGCTCTGCGGCTCGCTTAGGATCCCGGAAAGGAGACGCGGCGGAAAGCGGCCAG 720
QY 721 GCTGGGCCAGGATCTAGAAAGACTGCTGGCGCAGGCTCCCTGCCCGCGGCGCTGCTG 780
Db 721 GCTGGGCCAGGATCTAGAAAGACTGCTGGCGCAGGCTCCCTGCCCGCGGCGCTGCTG 780
QY 781 TCATGGACTCGTGGAGAGCTCGCTTCCCGCGGGACCTTCTCTGCAAGGCTCCACGTCCA 840
Db 781 TCATGGACTCGTGGAGAGCTCGCTTCCCGCGGGACCTTCTCTGCAAGGCTCCACGTCCA 840
QY 841 GGCACCGGCGCTCGGACACCCACCGCGCGGCGCACCTGCGCTGGGTGCGCCCTTAAC 900
Db 841 GGCACCGGCGCTCGGACACCCACCGCGCGGCGCACCTGCGCTGGGTGCGCCCTTAAC 900
QY 901 CCGGCGGTAGCTCGTTAAGATGGCGAAGTGTCCGGTCCGGAACACGCGAAACCCCAAT 960
Db 901 CCGGCGGTAGCTCGTTAAGATGGCGAAGTGTCCGGTCCGGAACACGCGAAACCCCAAT 960
QY 961 CCGGCTGCGCCGACCTCTCTGACCCCGCGGCCACGAGACTGGGCTCCCGACG 1020
Db 961 CCGGCTGCGCCGACCTCTCTGACCCCGCGGCCACGAGACTGGGCTCCCGACG 1020
QY 1021 CGGAGCGCTGCGCGGACACCGGTCGGTCCGAAACGAGGAGACTTTGTAACGCGAGTG 1080
Db 1021 CGGAGCGCTGCGCGGACACCGGTCGGTCCGAAACGAGGAGACTTTGTAACGCGAGTG 1080
QY 1081 TTTGCTCTTTTGAATAAACAAGATGCTTAACTGCTGAAAGCTCTGAAAGCTTGGCGCT 1140
Db 1081 TTTGCTCTTTTGAATAAACAAGATGCTTAACTGCTGAAAGCTCTGAAAGCTTGGCGCT 1140
QY 1141 AAAAGATGCTGGGTGACTTAGATGCTAGGATCAGTTTGTCTTCAATGTAAATGGACCAG 1200
Db 1141 AAAAGATGCTGGGTGACTTAGATGCTAGGATCAGTTTGTCTTCAATGTAAATGGACCAG 1200
QY 1201 CCGGACTCCGTAACGCACTAGCAGGGGACTGAAAGCGCTTTCAGGTACTGCTGGTGGC 1260
Db 1201 CCGGACTCCGTAACGCACTAGCAGGGGACTGAAAGCGCTTTCAGGTACTGCTGGTGGC 1260
QY 1261 GGTGATGCGCTACAGGCGGATCAGACAGTTTGTGTCTTCTGGAACCTTGACATGACCA 1320
Db 1261 GGTGATGCGCTACAGGCGGATCAGACAGTTTGTGTCTTCTGGAACCTTGACATGACCA 1320
QY 1321 CGGTAATGCTGAATCAGCAATATTTACAGATCAGACGCGATCATCTTCTTCAACATGA 1380
Db 1321 CGGTAATGCTGAATCAGCAATATTTACAGATCAGACGCGATCATCTTCTTCAACATGA 1380
QY 1381 TTTAAACACAGTTGACTTAATATATGCTGGATAAATGTAGAATACAAATATACCATCCAC 1440
Db 1381 TTTAAACACAGTTGACTTAATATATGCTGGATAAATGTAGAATACAAATATACCATCCAC 1440
QY 1441 CTGAGGCTTCTACTTCTGTAATTTTGGAGGCTGTTTAACTCTTGTACTCTAGCTTCT 1500
Db 1441 CTGAGGCTTCTACTTCTGTAATTTTGGAGGCTGTTTAACTCTTGTACTCTAGCTTCT 1500
QY 1501 TCATTAACAAAATAGGGGTACTAGCAGCGGGGTGCTCGCGCTGTAAATCCAGCACT 1560
Db 1501 TCATTAACAAAATAGGGGTACTAGCAGCGGGGTGCTCGCGCTGTAAATCCAGCACT 1560
QY 1561 TGGGAGGCGCGAGCGGCGGATCAGTTGAGGTGAGAAATTTTTCAGACGAGCTGGTCAAC 1620
Db 1561 TGGGAGGCGCGAGCGGCGGATCAGTTGAGGTGAGAAATTTTTCAGACGAGCTGGTCAAC 1620

```

QY 1621 ATGGGTGAACGCGCGCTCTACCAAAATATAAAATCTTAGCTGAGTGGTAGCGCATG 1680
DB 1621 ATGGGTGAACGCGCGCTCTACCAAAATATAAAATCTTAGCTGAGTGGTAGCGCATG 1680
QY 1681 ACTGTAATCCAGCAACTCAGGAGGCTGAGGAGAGAGAAATCGCTTGAACCTGGAGGCGGA 1740
DB 1681 ACTGTAATCCAGCAACTCAGGAGGCTGAGGAGAGAGAAATCGCTTGAACCTGGAGGCGGA 1740
QY 1741 GGTTCAGTGAGCTGAGTACCTACCTACCTACCTACCTACCTACCTACCTACCTACCTACCT 1800
DB 1741 GGTTCAGTGAGCTGAGTACCTACCTACCTACCTACCTACCTACCTACCTACCTACCTACCT 1800
QY 1801 GCCTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1860
DB 1801 GCCTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1860
QY 1861 GATGAGGGTTAAATTAAGTACACATACATAGCCCTAGCGAGTGGCTTATGCTGTAACTCT 1920
DB 1861 GATGAGGGTTAAATTAAGTACACATACATAGCCCTAGCGAGTGGCTTATGCTGTAACTCT 1920
QY 1921 CAACACTTTGGGAGTCTGTGGGCGGAGGATCACTTGAAGCCAGGAGTGGTGAAGTACTCT 1980
DB 1921 CAACACTTTGGGAGTCTGTGGGCGGAGGATCACTTGAAGCCAGGAGTGGTGAAGTACTCT 1980
QY 1981 GGGCAACAGAGACATCTCTATAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2040
DB 1981 GGGCAACAGAGACATCTCTATAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2040
QY 2041 GTGCACCTGCAGTCCAGCTACTAGGAGGCTGAGGTGGAGGACTGCTGAGCCAGCA 2100
DB 2041 GTGCACCTGCAGTCCAGCTACTAGGAGGCTGAGGTGGAGGACTGCTGAGCCAGCA 2100
QY 2101 GGTGAGGCTGAGTGAAGCATGATTTGTGCACTGCACTGCACTGCACTGCACTGCACTGCA 2160
DB 2101 GGTGAGGCTGAGTGAAGCATGATTTGTGCACTGCACTGCACTGCACTGCACTGCACTGCA 2160
QY 2161 GACCTTGTCTCAAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 2220
DB 2161 GACCTTGTCTCAAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 2220
QY 2221 CAATATGTGGTGGTGAATCTTCAATCTCTTCTGACTCAGATCTCATAGATTTTCTGA 2280
DB 2221 CAATATGTGGTGGTGAATCTTCAATCTCTTCTGACTCAGATCTCATAGATTTTCTGA 2280
QY 2281 ACTTTTGGAGAACTCTTGGCTCTCTGCATTTGCAACCGGTCAAAGGCACTCCCTTCTGCC 2340
DB 2281 ACTTTTGGAGAACTCTTGGCTCTCTGCATTTGCAACCGGTCAAAGGCACTCCCTTCTGCC 2340
QY 2341 ACCACAAAGCAATTTGAATTTTAACTTGAATATGCTCTCTCTCTCTCTCTCTCTCTCTCT 2400
DB 2341 ACCACAAAGCAATTTGAATTTTAACTTGAATATGCTCTCTCTCTCTCTCTCTCTCTCTCT 2400
QY 2401 TTAATCTTGGTCAGGGTCTCTCAACAGCCCTTTAGTCACTATGCCATTTGAATATGATGGC 2460
DB 2401 TTAATCTTGGTCAGGGTCTCTCAACAGCCCTTTAGTCACTATGCCATTTGAATATGATGGC 2460
QY 2461 CCTACAGCT 2469
DB 2461 CCTACAGCT 2469

```

RESULT 3

```

US-09-891-711-1
; Sequence 1, Application US/09891711
; Patent No. US20020082404A1
; GENERAL INFORMATION:
; APPLICANT: Pharmacia AB
; TITLE OF INVENTION: Promoter Sequences
; FILE REFERENCE: 00130
; CURRENT APPLICATION NUMBER: US/09/891,711
; CURRENT FILING DATE: 2001-06-26
; NUMBER OF SEQ ID NOS: 6

```

```

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 980
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-891-711-1

```

```

Query Match      39.7%; Score 980; DB 9; Length 980;
Best Local Similarity 100.0%; Pred. No. 6e-255;
Matches 980; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 353 CTCGCGGGCGAACAAGCGCTTGGGCACTTCCATTTGGGGCTGTTTTACTCCCAACTCTCGCGAG 412
DB 1 CTCGCGGGCGAACAAGCGCTTGGGCACTTCCATTTGGGGCTGTTTTACTCCCAACTCTCGCGAG 60
QY 413 ACTGGGGCGCGGGCGAGGAGCCCAAGCTGGGAGCCTCAGCTCCGCCGAGCAGCGCT 472
DB 61 ACTGGGGCGCGGGCGAGGAGCCCAAGCTGGGAGCCTCAGCTCCGCCGAGCAGCGCT 120
QY 473 GCCCTGTCTGCCCGCTCCCGGGCTTGGCTGGCGCTCTGGAGCGCGTGGGAGCGG 532
DB 121 GCCCTGTCTGCCCGCTCCCGGGCTTGGCTGGCGCTCTGGAGCGCGTGGGAGCGG 180
QY 533 GACCACGCGGAGGATGGAAGAGTGTCTCGCGACATTTGCGGGCGCGGGCGCGGTG 592
DB 181 GACCACGCGGAGGATGGAAGAGTGTCTCGCGACATTTGCGGGCGCGGGCGCGGTG 240
QY 593 CAGGCTGGAAGCGGAGGGGCGTGGCGAGGAGTGTGCGAGGCGCGAGAAACGGCTGGGG 652
DB 241 CAGGCTGGAAGCGGAGGGGCGTGGCGAGGAGTGTGCGAGGCGCGAGAAACGGCTGGGG 300
QY 653 AACCTTGTGTGGCTCTGGCGCTCTAGGATCCCGAAAGGAGGACGCGGCGGAA 712
DB 301 AACCTTGTGTGGCTCTGGCGCTCTAGGATCCCGAAAGGAGGACGCGGCGGAA 360
QY 713 GCGGCGAGGCTGGGCGAGGATCTAGAAAGCTTGGCGGCGAGGCTTCCCTGCCCCCGCG 772
DB 361 GCGGCGAGGCTGGGCGAGGATCTAGAAAGCTTGGCGGCGAGGCTTCCCTGCCCCCGCG 420
QY 773 GCTGTGTGTAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 832
DB 421 GCTGTGTGTAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 480
QY 833 CAGCTCCAGGACCGGCGCTGGGACACCCACCCCGCGCGGCGGCGGCGGCGGCTTGGGTGC 892
DB 481 CAGCTCCAGGACCGGCGCTGGGACACCCACCCCGCGCGGCGGCGGCGGCGGCTTGGGTGC 540
QY 893 CCTTAAACCGGCGGCTAGTCTGTTAAGATGGCAAGTGTCCGGTCCGGAAACACGCGAA 952
DB 541 CCTTAAACCGGCGGCTAGTCTGTTAAGATGGCAAGTGTCCGGTCCGGAAACACGCGAA 600
QY 953 CCCCATAATCCCGCTGCGCGACCTCTGACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1012
DB 601 CCCCATAATCCCGCTGCGCGACCTCTGACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
QY 1013 TCCGAGCGGCGAGCGGCTGCGGGGACCGGCTGGTGGGAAACGAGGAGCCTTGTCTAAC 1072
DB 661 TCCGAGCGGCGAGCGGCTGCGGGGACCGGCTGGTGGGAAACGAGGAGCCTTGTCTAAC 720
QY 1073 GCCACGTGTGTCTCTTTTGAATAAAGAAATGTTTAAACTGTCTGAAAAGCT 1132
DB 721 GCCACGTGTGTCTCTTTTGAATAAAGAAATGTTTAAACTGTCTGAAAAGCT 780
QY 1133 TCGCGCTTAAAGATGTCTGGGTGACTTAGATGTCTAGGATCACTTTGTTTCAATGATA 1192
DB 781 TCGCGCTTAAAGATGTCTGGGTGACTTAGATGTCTAGGATCACTTTGTTTCAATGATA 840
QY 1193 TGGACCGAGCCCGGACTTCCGTACGGCACTAGCAGGGGACTGAAAGCGCTTCAAGTACTGC 1252
DB 841 TGGACCGAGCCCGGACTTCCGTACGGCACTAGCAGGGGACTGAAAGCGCTTCAAGTACTGC 900
QY 1253 TGGTGGGCGGTGATGGCTTACAGGCGGATCAGACAGTGTGTGTCTTCTGGAACCTTGACA 1312

```



```
QY 1867 GGTAAATTA-----AGTACACACATAGCCCTAGCGCAGTGGCTTATGCCCTGTAA 1917
Db 259065 GCAAGCATTAATAATAATAATAATAATAAGCCAGGCACAGTGGCTCATAGCTGTAA 259066
QY 1918 TCTCAACACTTTGGGAGTCTGTGGCGGAGGATCACTTGGAGCCAGGAGTTTGAGACTAG 1977
Db 259005 TCCAGCACTTCGGGAGCGGAGGAGGAGGAGGATCACTTGGGTCAGGGGTTGAGACCAG 258946
QY 1978 TCTGGGCAACAGAGACATGCTCTTATAGTTGTGTTGTTGTTGTTGTTTACCAGGTGTGT 2037
Db 258945 CTTGGGCAACATGGTGAACCTCTCTCTACAAAAATAAAAAATTAGCCAGCGTGTGT 258886
QY 2038 GGTGTGCACCTGCAGTCCAGCTACTAGGGAGGCTGAGGTGGGAGCACTGCCTGAGCCCA 2097
Db 258885 GGTGCATGGCTATAATCCAGCTACTTGGGAGACTGAGGAGAGAAATCACTTTGAACCCG 258826
QY 2098 GAGGTTCGAGGCTGCAGTGAAGCATGATTTGTGCACCTGCACCTCCAGCTGGGCAACACAG 2157
Db 258825 AGAGATGAAGTTGCAAGTGAAGTGAAGTTGCGGCACTGCACCTCCAGCTGGGCAACAGAG 258766
QY 2158 CAAGACTTGTCTCAAAAAACAAACAAA 2185
Db 258765 TGAGACTTGTCAAAAAACAAACAAA 258738
RESULT 6
US-10-322-281-170/c
; Sequence 170, Application US/10322281
; Publication No. US20040126762A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer
; FILE REFERENCE: 529452001000
; CURRENT APPLICATION NUMBER: US/10/322,281
; CURRENT FILING DATE: 2002-12-17
; NUMBER OF SEQ ID NOS: 866
; SOFTWARE: PaetSeq for Windows Version 4.0
; SEQ ID NO 170
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(100445)
; OTHER INFORMATION: n = A,T,C or G
US-10-322-281-170
Query Match 13.4%; Score 330.8; DB 19; Length 100445;
Best Local Similarity 72.3%; Pred. No. 5.9e-78;
Matches 506; Conservative 0; Mismatches 172; Indels 22; Gaps 5;
QY 1494 AGCTTCTTCATTACAAAAATAGGGGTACTAGCCAGCGGGTGGCTCGCGCCCTGTAAATCC 1553
Db 47292 AGATTCTGTCTTAATAATAAAAAAGATAACGCGCGGGCGGTGTCTATGCCCTGTAAATCC 47233
QY 1554 CAGCACTTGGGAGCGGAGCGGAGCCCGGATCACTTGAAGTTCAGAGTTTCAGACCAAGCCT 1613
Db 47232 CAGCACTTGGGAGCGGAGCGGAGGTGGATCAC - GAGGTGAGGAGTTGAGACCATCCT 47175
QY 1614 GGTCAACATGGGTGAACCGCGGCTCTACAAAAATAATAAACTTAGCTGAGTGGTA 1673
Db 47174 GGTAAACAT - GGTGAACCGCGCTCTCTACTAAAAATAACAAAAATAGCTGGCGGTGGTG 47116
QY 1674 GCGCATGACTGTAAATCCAGCAACTCAGAGGCTGAGGCA - GAGAATCGCTTCAACTGG 1732
Db 47115 GCAGGTGCTTGTAGTCTCAGCTACTCAGGAGGCTGAGGAGGAGATGGGTGAACCCGG 47056
QY 1733 GAGCGGAGGTTGAGTGAAGTGAAGTTCGTAACCACTCCAGCTTCCAGCTTGGGCGACAGAGC 1792
Db 47055 GAGCGGAGGTTGAGTGAAGTGAAGTTCGTAACCACTCCAGCTTCCAGCTTGGGCGACAGAGC 46996
RESULT 7
US-10-723-860-2768/c
; Sequence 2768, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Marc S. Malandro
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer
; FILE REFERENCE: 529452001000
; CURRENT APPLICATION NUMBER: US/10/322,281
; CURRENT FILING DATE: 2002-12-17
; NUMBER OF SEQ ID NOS: 866
; SOFTWARE: PaetSeq for Windows Version 4.0
; SEQ ID NO 170
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(100445)
; OTHER INFORMATION: n = A,T,C or G
US-10-723-860-2768
Query Match 13.3%; Score 327.4; DB 20; Length 147620;
Best Local Similarity 71.0%; Pred. No. 5.9e-77;
Matches 489; Conservative 0; Mismatches 196; Indels 4; Gaps 4;
QY 1504 TTACAAAAATAGGGGTACTAGCCAGCGGGTGGCTCGCGCCCTGTAAATCCAGCACTTGG 1563
Db 107541 TCAAAAAATAATAATAATAGCGCGGGTGTGTGGCTCACGCTGTAAATCTCAGCACTTTG 107482
QY 1564 GGAGCGCGAGGCGAGCGGATCACTTGAAGTTCAGAGTTTTCAGACCAAGCTGGTCAACATG 1623
Db 107481 GGAGCGCGAGGCGCGGAGTCACTTGAAGTTCAGAGTTTTCAGACCAAGCTGGTCAACAT - 107423
QY 1624 GGTGAACCGCGCTCTACAAAAATAATAAACTTAGCTGAGTGTGGTAGCGCATGACT 1683
Db 107422 GGTGAACCGCGCTCTCTACTAAAAATAACAAAA - TTAGTGGCGGTGGTGGTGGCT 107364
QY 1684 GTAATCCAGCAACTCAGAGGCTGAGGCA - GAGATCCGCTTGAACCTGGGAGCGGAGG 1742
Db 107363 CTAATCCAGCACTTGGGAGGCTGAGGCAAGAAATCAATTAATCCAGGAGCGGAGG 107304
QY 1743 TTGCAGTGAAGTGAAGTTCGTAACCACTCCAGCTTCCAGCTTGGGCGACAGAGCACTCTGC 1802
```


1568 GCCGAGCCGCGATCACTTGGTTCAGAGTTTCAGACCCAGCTGCTCAACATCGGTC 1627
 Db 88688 GCCGAGCCGCGCGATCACTTGGTTCAGAGTTTCAGACCCAGCTGCTCAACATCGGTC 88632
 1628 AAACGGCGGCTCTACCAAAAATAATAAACTTACCTGAGTGTGGTACGCGCATGACTGTAA 1687
 Db 88631 AAACCGGCTCTACTTAAATAACAATAATAGCCGGGCTAGTGTGGCGGCGCTGTAG 88572
 1688 TCCAGCAACTCAGGAGGCTGAGGCA - GAGAAATCGCTTGAACCTGGAGGCGGAGGTTGC 1746
 Db 88571 TCCAGCACTTGGAGGCTGAGGCAAGGAGATGGCTGAAACCGGAGGCGGAGCTTGC 88512
 1747 AGTGAGCTGAGATCGTACCACTGCACTCCAGCTTGGGCGACAGAGCGAGACTTGCCTTA 1806
 Db 88511 AGTGAGCGGAGATTGGCGCACTGCACTCCAGCTTGGGCGACAGAGCGAGACTCCGCTCA 88452
 1807 AAAATAAAATAAATTTTAAAAAATAA - - - - - GGGGTACTAATATCTACCTTAAAG 1860
 Db 88451 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAATGGAATTCCTTTAACTGATTTGGGT 88392
 1861 GATGAGGGTTAAATTAAGTACACATAAGCCCTAGCGAGTGGCTTATGCTTGTAACTGTAACT 1920
 Db 88391 GACAAACGACAGCAAAATTAAGAAATATAGGCCAGGTGCGGTAGCTCATGCTGTAAATGC 88332
 1921 CAACACTTTGGGAGTCTGTGGCGGAGGATCACTTGGCCAGGAGTTTGAGACTAGTCT 1980
 Db 88331 CAGCACTTTGGGAGCCGAGCGGAGTGGATCACTGAGCGCAGGAGTTTGAGACCAAGCCT 88272
 1981 GGGCAACAGAGACATGCTCTATAGTTGTGTTGGTTTGTGTTTGTGTTTGTGTTTGTGTTT 2040
 Db 88271 GGGCAACATGTTGAAACCCGCTCTACTTAAATAACAATAATAGTGGCAATTTGGC 88212
 2041 GTGCACCTGCAGTCCAGCTTACTAGGAGGCTGAGGTGGAGGACTGCTTGGAGCCAGGA 2100
 Db 88211 ACGGCTGTAGTCCAGCTACTCAAGAGGCTGAGGAGGAGAAATCACTTAAACCCAGGA 88152
 2101 GGTGAGGCTGAGTGGAGGATGTTGCACTGCACTCCAGCTGCGGCAACAGCA 2160
 Db 88151 GATGAGGTTTGGGTGAGCAGGATCATGCTTGCACCTCCAGCTGCGGCGGCGGCAAC 88092
 2161 GACCTTGTCTCAAAAACAAACAAAAA 2186
 Db 88091 GACTCCATCTCAAAAATAAATAATA 88066
 RESULT 10
 ; US-10-737-082-70/c
 ; Sequence 70, Application US/10737082
 ; Publication No. US20050130170A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Bayer Healthcare LLC
 ; APPLICANT: Beard, Chris
 ; APPLICANT: Burgess, Chris
 ; APPLICANT: Gannon, Allison
 ; APPLICANT: Harvey, Jeanne
 ; APPLICANT: Lechner, John F.
 ; APPLICANT: Li, Zheng
 ; TITLE OF INVENTION: Identification and Verification of Methylation Marker Sequences
 ; FILE REFERENCE: 1657/2032
 ; CURRENT APPLICATION NUMBER: US/10/737,082
 ; CURRENT FILING DATE: 2003-12-16
 ; PRIOR FILING DATE: 2003-12-16
 ; NUMBER OF SEQ ID NOS: 300
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 70
 ; LENGTH: 354592
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-10-737-082-70

Best Local Similarity 71.0%; Pred. No. 9.4e-76;
 Matches 487; Conservative 0; Mismatches 189; Indels 10; Gaps 4;
 QY 1508 AAAAAATAGGGGTACTAGCCAGCGGGGTGGCTGGCGCTGTAAATCCAGCAGCTTGGGGAG 1567
 Db 91139 AAGAAATGAAATTGGCGCGGGCGGGTGGCTCAGCGCTGTAAATCCAGCAGCTTGGGGAG 91080
 QY 1568 GCGGAGGAGCGGGATCACTTGGGTGAGGTCAGAGTTTTCAGACCCAGCTGCTCAACATCGGTC 1627
 Db 91079 GCGGAGGCGGGCGGATCAC - - GAGGTGAGGATCGAGACCATCCCGG - CTAAAAACGGTG 91023
 QY 1628 AAAACGCGGCTCTACCAAAAATAATAAACTTACCTGAGTGTGGTACGCGCATGACTGTAA 1687
 Db 91022 AAACCCGCTCTACTTAAATAACAATAATAGCGGGCTAGTGGCGGCGCTGTAG 90963
 QY 1688 TCCAGCAACTCAGGAGGCTGAGGCA - GAGAAATCGCTTGAACCTGGAGGCGGAGGTTGC 1746
 Db 90962 TCCAGCTACTTGGGAGGCTGAGGCGAGGAGAAATGGCTGAAACCGGAGGCGGAGCTTGC 90903
 QY 1747 AGTGAGCTGAGATCGTACCACTGCACTCCAGCTTGGGCGACAGAGCGAGACTTGCCTTA 1806
 Db 90902 AGTGAGCGGAGATTGGCGCACTGCACTCCAGCTTGGGCGACAGAGCGAGACTCCGCTCA 90843
 QY 1807 AAAATAAAATAAATTTTAAAAAATAA - - - - - GGGGTACTAATATCTACCTTAAAG 1860
 Db 90842 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAATGGAATTCCTTTAACTGATTTGGGT 90783
 QY 1861 GATGAGGGTTAAATTAAGTACACATAAGCCCTAGCGAGTGGCTTATGCTTGTAACTGTAACT 1920
 Db 90782 GACAAACGACAGCAAAATTAAGAAATATAGGCCAGGTGCGGTAGCTCATGCTGTAAATGC 90723
 QY 1921 CAACACTTTGGGAGTCTGTGGCGGAGGATCACTTGGCCAGGAGTTTGAGACTAGTCT 1980
 Db 90722 CAGCACTTTGGGAGCCGAGCGGAGTGGATCACTGAGCGCAGGAGTTTGAGACCAAGCCT 90663
 QY 1981 GGGCAACAGAGACATGCTCTATAGTTGTGTTGGTTTGTGTTTGTGTTTGTGTTTGTGTTT 2040
 Db 90662 GGGCAACATGTTGAAACCCGCTCTACTTAAATAACAATAATAGTGGCAATTTGGC 90603
 QY 2041 GTGCACCTGCAGTCCAGCTTACTAGGAGGCTGAGGTGGAGGACTGCTTGGAGCCAGGA 2100
 Db 90602 ACGGCTGTAGTCCAGCTACTCAAGAGGCTGAGGAGGAGAAATCACTTAAACCCAGGA 90543
 QY 2101 GGTGAGGCTGAGTGGAGGATGTTGCACTGCACTCCAGCTGCGGCAACAGCA 2160
 Db 90542 GATGAGGTTTGGGTGAGCAGGATCATGCTTGCACCTCCAGCTGCGGCGGCGGCAAC 90483
 QY 2161 GACCTTGTCTCAAAAACAAACAAAAA 2186
 Db 90482 GACTCCATCTCAAAAATAAATAATA 90457
 RESULT 11
 ; US-10-765-790-70/c
 ; Sequence 70, Application US/10765790
 ; Publication No. US20050130172A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Bayer Healthcare LLC
 ; APPLICANT: Beard, Chris
 ; APPLICANT: Burgess, Chris
 ; APPLICANT: Gannon, Allison
 ; APPLICANT: Harvey, Jeanne
 ; APPLICANT: Lechner, John F.
 ; APPLICANT: Li, Zheng
 ; TITLE OF INVENTION: Identification and Verification of Methylation Marker Sequences
 ; FILE REFERENCE: 1657/2035
 ; CURRENT APPLICATION NUMBER: US/10/765,790
 ; CURRENT FILING DATE: 2004-01-27
 ; PRIOR FILING DATE: 2003-12-16
 ; NUMBER OF SEQ ID NOS: 300
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 70


```

; LENGTH: 354592
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-765-790-70

Query Match      13.1%; Score 323.6; DB 22; Length 354592;
Best Local Similarity 71.0%; Pred. No. 9.4e-76;
Matches 487; Conservative 0; Mismatches 189; Indels 10; Gaps 4;

QY 1508 AAAAATAGGGGTACTAGCCAGCGGGGTGGCTCGCGCTGTAAATCCACAGCTTGGGGAG 1567
Db 91139 AAGAAATGGAATTTGCGCGCGGGCGGTGCTCACGCTGTAAATCCACAGCTTGGGGAG 91080
QY 1568 GCGAGGCGAGCGGATCACTTGAGGTTCAGAGTTTCAGACACGCTGTGTCAACATGGGTG 1627
Db 91079 GCGAGGCGGCGGATCAC--GAGGTTCAGAGATCGAGACCATCCCGG--CTAAACGGGTG 91023
QY 1628 AAACGCGGCTCTACCAAAATATAAAAACTTAGCTGAGTGTGGTAGCGCATGACTGTAA 1687
Db 91022 AAACCCGCTCTACTATAAAATACAAAATTTAGCCGGGCTAGTGGCGGCGCTGTAG 90963
QY 1688 TCCAGCACTCAGGAGGCTGAGGCA--GAGATCGCTTGAACCTGGAGCGCGAGGTTCG 1746
Db 90962 TCCAGCTACTTTGGGAGGCTGAGCGAGGAGATGGCGTGAACCCGGAGGCGGAGCTTCG 90903
QY 1747 AGTGAGCTGAGATCGTACCACCTGCTCCAGCTTTGGCGGACAGAGAGAGACTCTGCTTA 1806
Db 90902 AGTGAGCGGAGATTTGGCGCACTGCATCTCCAGCTTGGCGGACAGAGAGACTCCGCTCA 90843
QY 1807 AAAATAAATAAATAAATTTTAAAAAAAATA-----GGGCTACTAATAATCTACCTTAAAG 1860
Db 90842 AAAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 90783
QY 1861 GATGAGGTTAAATTAAGTACACATAAGCCCTAGCGAGTGGCTTATGCTGTAAATCT 1920
Db 90782 GACAAACGACAGCAAAATTAAGAATAATTAGCGCAGGTGGCGTAGCTCATGCTGTAAATGC 90723
QY 1921 CAACACTTTGGAGTCTGTGGCGGAGGATCACTTGAGCCAGGAGTTTGAGACTAGTCT 1980
Db 90722 CAGCACTTTGGAGCGCGAGCGGATGATCACTGAGGCGAGGATTTGAGACACGCT 90663
QY 1981 GGGCAACAGACATGCTCTATAGTTGTGTTGGTTTGTGTTTGTGTTTGTGTTTGTGTTG 2040
Db 90662 GGGCAACATGTTGAACCCGCTCTCTACTAAAAATACAAAATAGCTGGGCATTTGGC 90603
QY 2041 GTCCACTGCGAGTCCAGCTACTAGGAGCTGAGGTGGAGAGCTGCTGAGCCCGAGGA 2100
Db 90602 ACGGCTGTAGTCCAGCTACTCAAGAGGCTGAGGCGAGGAGATCACTTAAACCCAGGA 90543
QY 2101 GGTGAGGCTGAGTGAGCATGTTGTCCACTGCACTCCAGCTGGGCGAACAACAGCAA 2160
Db 90542 GATGGAGTTGGCGGTGAGCAGGATCAATGCTCTGCACTCCAGCTGGGCGAGTGAACAA 90483
QY 2161 GACCTGCTCAAAAAACAAACAAAA 2186
Db 90482 GACTCCATCTCAAAAAATAAATAA 90457

RESULT 12
US-10-719-993-6815/c
; Sequence 6815, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6815
; LENGTH: 1980090

```

```

; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1980090)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-2)
US-10-719-993-6815

Query Match      13.1%; Score 322.6; DB 20; Length 1980090;
Best Local Similarity 71.6%; Pred. No. 3.8e-75;
Matches 489; Conservative 4; Mismatches 178; Indels 12; Gaps 5;

QY 1506 AAAAAATAGGGGTACTAGCCAGCGGGGTGGCTCGCGCTGTAAATCCACAGCTTGGGG 1565
Db 1405235 ACAGATTATACAGATGACGCGCGGCGAGCTGCTCAGCGCTGTAAATCCACAGCTTGGC 1405176
QY 1566 AGCGGAGCGAGCGCGGATCACTTGAGGTTCAGAGTTTCAGACGAGCTGTGTCACATGGG 1625
Db 1405175 AGCGGAGCGGCGGCGGATCAC--GAGGTTCAGAGATTCAGACCATCTGCTTAACA-AG 1405119
QY 1626 TGAACGCGCGCTCTACCAAAATATAAAAACTTAGCTGAGTGTGGTAGCGCATGACTGT 1685
Db 1405118 AGAAACYCGTCTCTACTAAATATACAAAATTTAGCYGGCGTGGGCGCGCGCTCT 1405059
QY 1686 AATCCAGCACTCAGGAGGCTGAGGCA--GAGATCGCTTGAACCTGGAGCGCGAGGTT 1744
Db 1405058 AGTCCAGCTACTCGGAGGCTGAGCGAGGAGATGGCGTGAACCCGGGAGCGGAGCTT 1404999
QY 1745 GCAGTGAGCTGAGATCGTACCACTGCATCCAGCTTGGCGACAGAGCGAGACTCTGCCT 1804
Db 1404998 GCAGTGAGCGGAGATGGCACCTGCATCCAGCTTGGCGACAGAGCGAGACTCTGTCT 1404939
QY 1805 TAAAAATAAATAAATAAATTTTAAAAAATAAAGGGTACTAATAATCTACCTTAAAGGATG 1864
Db 1404938 CAAAAAATAAATAAATAAATTTTACAGATGACAACTGTGGATGCAGCTAC-----GTAG 1404886
QY 1865 AGGGTTAAATTAAGTACACATAAGCCCTAGCGAGTGGCTTATGCTGTAAATCTCAAC 1924
Db 1404885 TGTGTAGAAAGAAATGTATAAACTTTGGCGGCGCAGTGGCTCAVCCCGTAAATCCCAAC 1404826
QY 1925 ACTTTGGAGTCTGTGGCGGAGGATCACTTGAGCCCGAGGATTTGAGACTAGTCTGGGC 1984
Db 1404825 ATTCTGGAGGCTGAGGAGGCGGATCACTGAGGTTCAGAGTTTCAGACCGCTTAGCC 1404766
QY 1985 AACA--GAGACATGCTCTATAGTTGTGTTGGTTTGTGTTTGTGTTTGTGTTGTTG 2043
Db 1404765 AACATGGCAAAACCTCTCTACTAAAAATAAATAAATAAATAAATAAATAAATAAATAA 1404706
QY 2044 CACCTGAGTCCAGCTACTAGGAGGCTGAGGTGGAGGACTGCTGAGCCAGGAGGT 2103
Db 1404705 TGCTGTAAATCCAGCTGCTCGGAGGCTGAGGCGAGGAGATCACTTGAACCCAGGAGGC 1404646
QY 2104 CGAGGTGAGTGGAGCATGATTTGCCACTGCACTCCAGCTTGGCGAACAACAGCAAGAC 2163
Db 1404645 GGAGGTTGCAGTGAGCGGAGATAGCCCACTGCCTCCAGCTTGGCGAGAGCGAGAC 1404586
QY 2164 CTGTCTCAAAACAAACAAAA 2186
Db 1404585 TCGGTCTCAAAACAAACAAAA 1404563

RESULT 13
US-10-741-600-17676/c
; Sequence 17676, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001499
; CURRENT APPLICATION NUMBER: US/10/741,600
; NUMBER OF SEQ ID NOS: 2003-12-22
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 73997

```


Db 68222 AAAATAAAGGTTTTTTTAAA 68203
Search completed: August 12, 2005, 08:55:00
Job time : 1574 secs

APPLICANT: Morris, David W.
APPLICANT: Engelhard, Eric K.
TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
TITLE OF INVENTION: CANCER
FILE REFERENCE: 529452000122
CURRENT APPLICATION NUMBER: US/10/087,192
CURRENT FILING DATE: 2002-03-01
PRIOR APPLICATION NUMBER: US 09/747,377
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 09/798,586
PRIOR FILING DATE: 2001-03-02
NUMBER OF SEQ ID NOS: 2059
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 844
LENGTH: 91760
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)..(91760)
OTHER INFORMATION: n = A, T, C or G
US-10-087-192-844

Query Match		13.0%	Score 321.6;	DB 13;	Length 91760;
Best Local Similarity		72.1%	Pred. No. 1.8e-75;		
Matches 490;		Conservative 0;	Mismatches 179;	Indels 11;	Gaps 5;
Qy	1523	AGCCAGCGGGTGGCTGGCGCTGTAAATCCAGCACTTGGGAGGCGCGAGGCGCGGA	1582		
Db	68876	AGCGGGCGGAGTGGCTCAAGCTGTAAATCCAGCACTTGGGAGGCGCGAGGCGGA	68817		
Qy	1583	TCACCTTGAGGTCAGAGCTTTTCAGACCACTGGTCAACATGGGTGAAACCGCGCTCTAC	1642		
Db	68816	TCAC--GAGGTCAGAGCTTTGAGACCACTGGTCAACAT--GGTGAACCCCGCTCTCTAC	68760		
Qy	1643	CAAAAATATAAAATTTAGCTGAGTGTGGTAGCGCATGCTGTAAATCCAGCAACTCAGG	1702		
Db	68759	TAAATATACAAAATTAGCCAG---GCACGGTGGCAGCGCTTAACCCCAACTACTCCGG	68703		
Qy	1703	AGGCTGAGGC--AGAGATCGCTTGAACCTGGGAGGCGGAGGTTGCAGTGAGCTGAGATCG	1761		
Db	68702	AGGCTGATGCGGGAGAATCACTTGAACCGAGGAGGTGAAGGTTGCAGTGAGCCAGACCG	68643		
Qy	1762	TACCACCTGCACCTCCAGCTTGGCGCACAGAGAGACTCTGCCTTAAAAATAAATAAATAA	1821		
Db	68642	TGCCATTGCACCTCCAGCTTGGCGCACAGAGAGACTCTGTCTCAAAAAATAAATAAATAA	68583		
Qy	1822	TTTTTAAAAAATAGGGGTACTAATATCTACCTTAAAGGATGAGGGTTAAATTAAGTAC	1881		
Db	68582	AAAAAATAAAGCAAGGTCTGNAACAGTGACATAGTATGCCACGTTTGTGTTTTAAGA	68523		
Qy	1882	ACACATAGCCCTAGCGCAGTGGCTTATGCTGTAACTCAACACTTTGGGAGTCTGTGG	1941		
Db	68522	AAAGTGTAGGCTGGATGCAATGGCTCGCACCTACAGTCCAGACACTTTGGGAGCGGAGG	68463		
Qy	1942	CGGAGGATCACTTGAGCCCAAGGAGTTTGAGACTAGTCTGGGCAACA----GAGACATGT	1997		
Db	68462	CAGAAGGATTCTTTGAGCCAGGAGTTCAAGACCAGCCTGGGCAACATAGCGAAGCCTG	68403		
Qy	1998	CTCTATAGTGTGTTTGTGTTTTTTTACAGGTTGGTGTGTGCACCTGTCAGTCCCA	2057		
Db	68402	TTCTATGAAGATAAATAAATAATTAAGCCAGGCTGGTGTGTGCACCTGTAGTCCCA	68343		
Qy	2058	GCTACTAGGAGGCTGAGGTGGGAGGACTGCTGAGCCAGAGGTCGAGGCTGAGTGA	2117		
Db	68342	GCTACTTGGGAGGCTGTGGTAGGAGGATGGCTTGAGTCTAGGAGTTTGAGGCTGCAGTGA	68283		
Qy	2118	GCCATGATGGCCACTGCACTCCAGCCTGGGCAACAGAGACCTTGTCTCAAAAAC	2177		
Db	68282	GCTATGATGGGCGCGCTGCACTCCAGCCTGGGCAACACAGTGAGACCTGTCTCTAAAT	68223		
Qy	2178	AAACAAAAAGCATACTCATATA	2197		

This Page Blank (uspto)

~~This Page Blank (uspto)~~
~~This Page Blank (uspto)~~

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 11, 2005, 22:55:02 ; Search time 411 Seconds
(without alignments)
9829.604 Million cell updates/sec

Title: US-10-829-118-2

Perfect score: 2469

Sequence: 1 ttgagtcgtctggaggctc.....gaatacatggccctacagct 2469

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA:*

- 1: /cgn2_6/prodata/1/ina/5A COMB.seq.*
- 2: /cgn2_6/prodata/1/ina/5B COMB.seq.*
- 3: /cgn2_6/prodata/1/ina/6A COMB.seq.*
- 4: /cgn2_6/prodata/1/ina/6B COMB.seq.*
- 5: /cgn2_6/prodata/1/ina/PCFUS COMB.seq.*
- 6: /cgn2_6/prodata/1/ina/backfileseq.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2469	100.0	2469	4	US-09-891-711-2
C 2	2452.2	99.3	67156	4	US-09-949-016-12284
C 3	2452.2	99.3	67157	4	US-09-949-016-16558
4	980	39.7	980	4	US-09-891-711-1
5	346	14.0	32010	4	US-09-949-016-13127
C 6	340.4	13.8	40655	4	US-09-949-016-12032
C 7	340.4	13.8	40655	4	US-09-949-016-15919
C 8	332	13.4	187136	4	US-09-949-016-17231
C 9	330.8	13.4	84425	4	US-09-949-016-17402
10	324.2	13.1	111454	4	US-09-949-016-14196
11	324.2	13.1	111459	4	US-09-949-016-12337
C 12	322.6	13.1	43255	4	US-09-949-016-11909
C 13	322.2	13.0	112219	4	US-09-949-016-12453
C 14	322.2	13.0	112222	4	US-09-949-016-14324
C 15	322.2	13.0	113186	4	US-09-949-016-17572
C 16	318	12.9	29067	4	US-09-949-016-12535
C 17	318	12.9	29093	4	US-09-949-016-17476
C 18	317.8	12.9	63760	4	US-09-949-016-14087
C 19	317.8	12.9	63760	4	US-09-949-016-14088
C 20	317	12.8	84587	4	US-09-949-016-15733
C 21	316.8	12.8	89843	4	US-09-949-016-13736
C 22	316.8	12.8	89844	4	US-09-949-016-13656
C 23	316	12.8	63783	4	US-09-949-016-13576
C 24	315.8	12.8	43657	4	US-09-949-016-13777
C 25	314.8	12.8	87562	4	US-09-949-016-13685
C 26	314.2	12.7	24166	4	US-09-949-016-13755
27	314.2	12.7	91665	4	US-09-949-016-12234

ALIGNMENTS

RESULT 1

US-09-891-711-2
; Sequence 2, Application US/09891711
; Patent No. 6723553
; GENERAL INFORMATION:
; APPLICANT: Pharmacia AB
; TITLE OF INVENTION: Promoter Sequences
; FILE REFERENCE: 00130
; CURRENT APPLICATION NUMBER: US/09/891,711
; CURRENT FILING DATE: 2001-06-26
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2469
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-891-711-2

Query Match 100.0%; Score 2469; DB 4; Length 2469;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2469; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	TTGAGTCTCTGAGGCTCCGGCCAGAGCAGGCGGTATTGTTTTCACCTCGGTGAATGCT	60
DB	1	TTGAGTCTCTGAGGCTCCGGCCAGAGCAGGCGGTATTGTTTTCACCTCGGTGAATGCT	60
QY	61	CATTTTCACTAAAGAAACACAGGCAACGGAACAAGCTCGGAGCGCGCAGACCCCGCA	120
DB	61	CATTTTCACTAAAGAAACACAGGCAACGGAACAAGCTCGGAGCGCGCAGACCCCGCA	120
QY	121	GGGCCCGGTACAGGCACTGTGTGTCTCAAAACAGCCCGGAGCCCGGCGCCACCTCCC	180
DB	121	GGGCCCGGTACAGGCACTGTGTGTCTCAAAACAGCCCGGAGCCCGGCGCCACCTCCC	180
QY	181	CGACCCCGCCCGCCCGCCCGCCCGCTCGCTCGGGGCTCGGACCAACCGGACACCT	240
DB	181	CGACCCCGCCCGCCCGCCCGCCCGCTCGCTCGGGGCTCGGACCAACCGGACACCT	240
QY	241	GAGCAGCGGGCGCCGACACCGCTAGGCGGAGCGGGTCTGGGAGCCCGCGCGGGCGGCTG	300
DB	241	GAGCAGCGGGCGCCGACACCGCTAGGCGGAGCGGGTCTGGGAGCCCGCGCGGGCGGCTG	300
QY	301	ACGTACTCGCCCGCGGAGCTCAGGCGCGGGCGCGCGGATGACGGGCGCTTCGCGG	360
DB	301	ACGTACTCGCCCGCGGAGCTCAGGCGCGGGCGCGCGGATGACGGGCGCTTCGCGG	360
QY	361	CGACACCGCTGGGCACTCCATTCGGGGCTGTTTACTCCCACTCTCGCGGAGACTGGCG	420
DB	361	CGACACCGCTGGGCACTCCATTCGGGGCTGTTTACTCCCACTCTCGCGGAGACTGGCG	420

421 GC CGG C C A G C A G G C C C A C A G C T G G G A G C C T C A G C T C C G C G A C C C A G C T G C C C T G T C 480
421 G C C G G C C A G C A G G C C C A C A G C T G G A G C C T C A G C T C C G C G A C C C A G C T G C C C T G T C 480
481 T G T C C C G C G C T C C C G G G C T T G C G T C G C G C T C T G A C G C C G T G G G A C C A C G C 540
481 T G T C C C G C G C T C C C G G G C T T G C G T C G C G C T C T G A C G C C G T G G G A C C A C G C 540
541 C G G A G A T G A C A A A G T G C T C G C A C A T T T T G C G C G C G G G C C G T G G C A G G T G G 600
541 C G G A G A T G A C A A A G T G C T C G C A C A T T T T G C G C G C G G G C C G T G G C A G G T G G 600
601 A A G C G A G G G C G T G C C A G C A G C T G C C A G C G G C A G A A C G C G T G G G G A A C C C T T G 660
601 A A G C G A G G G C G T G C C A G C A G C T G C C A G C G G C A G A A C G C G T G G G G A A C C C T T G 660
661 G T C C G C T C T G C G C G T C G C T T A G A T C C C C G A A A G A G A C A G G C G C G A A A G C G C C A G 720
661 G T C C G C T C T G C G C G T C G C T T A G A T C C C C G A A A G A G A C A G G C G C G A A A G C G C C A G 720
721 G C T G G C C A G G A T C T A G A A A G A C T G C T G C G C A G G C T C C T G C C C C C G G G C C T G C T G 780
721 G C T G G C C A G G A T C T A G A A A G A C T G C T G C G C A G G C T C C T G C C C C C G G G C C T G C T G 780
781 T C A T G A C T C G T G A G A G C T C G C T T C C C G C G G A C C C T T C T G C A G G G T C C A G C T C C A 840
781 T C A T G A C T C G T G A G A G C T C G C T T C C C G C G G A C C C T T C T G C A G G G T C C A G C T C C A 840
841 G G C A C C G G G C T C G G A C A C C C C A C C C C G G C G G G C A C C T G C C C T G G G T G C C C T T A A C 900
841 G G C A C C G G G C T C G G A C A C C C C A C C C C G G C G G G C A C C T G C C C T G G G T G C C C T T A A C 900
901 C C G G G C G T A G C T G T T A G A T G G C A A G T G C C G T C C G G A C A C G C G A A C C C A A T 960
901 C C G G G C G T A G C T G T T A G A T G G C A A G T G C C G T C C G G A C A C G C G A A C C C A A T 960
961 C C C G C T G C C C A C C T C T G A C C C C G C C C A C G G A C A C A G A C T G G G C C T C C C A G C 1020
961 C C C G C T G C C C A C C T C T G A C C C C G C C C A C G G A C A C A G A C T G G G C C T C C C A G C 1020
1021 C G C A G C G C T G C G G G A C A C C C G T G C T G C G A A C G G A G A C C T T T G A A C C C A C G T G 1080
1021 C G C A G C G C T G C G G G A C A C C C G T G C T G C G A A C G G A G A C C T T T G A A C C C A C G T G 1080
1081 T T T G C T C T T T T G A A A A A C A A G A T A A T G T T T A A C T G T C T G A A A G C T T G C G C C T 1140
1081 T T T G C T C T T T T G A A A A A C A A G A T A A T G T T A A C T G T C T G A A A G C T T G C G C C T 1140
1141 A A A G A T G T C G G T G A C T T A G A T G T A G A T C A G T T G T T T C A A T G T A A T G G A C C A G 1200
1141 A A A G A T G T C G G T G A C T T A G A T G T A G A T C A G T T G T T T C A A T G T A A T G G A C C A G 1200
1201 C C C G A C T C C G T A C G G C A C T A G C A G G G A C T G A A A G C T T C A G G T A C T G C T G G G C 1260
1201 C C C G A C T C C G T A C G G C A C T A G C A G G G A C T G A A A G C T T C A G G T A C T G C T G G G C 1260
1261 G G T A T G C G T A C A G C C G A T C A G A C A G T T T G T C T T C T G A A C T T G A C A C T G A C C A 1320
1261 G G T A T G C G T A C A G C C G A T C A G A C A G T T T G T C T T C T G A A C T T G A C A C T G A C C A 1320
1321 C G G T A A T G C T G A A C C A A T T A C A G A T C A C A G C G A T C A T C T T C C T C A A C A T G A 1380
1321 C G G T A A T G C T G A A C C A A T T A C A G A T C A C A G C G A T C A T C T T C C T C A A C A T G A 1380
1381 T T T A A C A C A G T T G A C T T A A T T A G T G G A T A A A T G T A G A A T C A C A A A T T A C C A T A C C C C A C 1440
1381 T T T A A C A C A G T T G A C T T A A T T A G T G G A T A A A T G T A G A A T C A C A A A T T A C C A T A C C C C A C 1440
1441 C T C A G G C T T C A C T T G C T A A T T T T G A C A G G T T G T T T A A C C T C T T T G T A C C T C A G C T T C T 1500
1441 C T C A G G C T T C A C T T G C T A A T T T T G A C A G G T T G T T T A A C C T C T T T T A C C T C T T T G T A C C T C A G C T T C T 1500
1501 T C A T T A C A A A A T A G G G T A C T A G C C A G G G G T G C T C G C C T G T A A T C C A G C A C T 1560

1501 T C A T T A C A A A A T A G G G T A C T A G C A G C G G G T G C T C G C C T G T A A T C C C A G C A C T 1560
1561 T G G G A G C C G A G G C A G C C G G A T C A C T T G A G T C A A A G T T T C A G A C C A G C C T G T C A A C 1620
1561 T G G G A G C C G A G G C A G C C G G A T C A C T T G A G T C A A A G T T T C A G A C C A G C C T G T C A A C 1620
1621 A T G G G T G A A A C C C G G C T C T A C A A A A A T A T A A A A C T T T A G C T G A G T G G T A G C G C A T G 1680
1621 A T G G G T G A A A C C C G G C T C T A C A A A A A T A T A A A A C T T T A G C T G A G T G G T A G C G C A T G 1680
1681 A C T G T A A T C C C A G C A A C T C A G G A G C T G A G G C A G A A A T C G C T T G A A C C T G G A G C G G A 1740
1681 A C T G T A A T C C C A G C A A C T C A G G A G C T G A G G C A G A A A T C G C T T G A A C C T G G A G C G G A 1740
1741 G G T G C A G T G A C T G A G A T C G T A C A C T C G A C T C C A G C T T G G G C G A C A G A G C A G A C T C T 1800
1741 G G T G C A G T G A C T G A G A T C G T A C C A C T G C A C T C C A G C T T G G G C G A C A G A G C A G A C T C T 1800
1801 G C C T T A A A A A T A A A T A A T A A T T T T A A A A A A T A G G G T A C T A A T A T C T A C C T T A A A G 1860
1801 G C C T T A A A A A T A A A T A A T A A T T T T A A A A A A T A G G G T A C T A A T A T C T A C C T T A A A G 1860
1861 G A T G A G G T T A A A T T A A G T A C A C A T A A G C C C T A G C C A G T G G C T T A T G C C T G T A A T C T 1920
1861 G A T G A G G T T A A A T T A A G T A C A C A T A A G C C C T A G C C A G T G G C T T A T G C C T G T A A T C T 1920
1921 C A A C A C T T T G G A G T C T G T G C C A G G A G A T C A C T T G A G C C C A G A G A G T T T G A C A C T A G T C T 1980
1921 C A A C A C T T T G G A G T C T G T G C C A G G A G A T C A C T T G A G C C C A G A G A G T T T G A C A C T A G T C T 1980
1981 G G G C A A C A G A C A T G T C T A T A G T T G T T T G T T T T T A C C A G G T G T G T G T 2040
1981 G G G C A A C A G A C A T G T C T A T A G T T G T T T G T T T T T A C C A G G T G T G T G T 2040
2041 G T G C A C C T G C A G T C C C A G C T A C T A G G A G G C T G A G T G G G A G A C T G C C T G A G C C C A G G A 2100
2041 G T G C A C C T G C A G T C C C A G C T A C T A G G A G G C T G A G T G G G A G A C T G C C T G A G C C C A G G A 2100
2101 G G T G A G G C T G A G T G A G C C A T G A T T G T G C C A C T G C A C T C C A G C C T G G G C A A C A C A G C A A 2160
2101 G G T G A G G C T G A G T G A G C C A T G A T T G T G C C A C T G C A C T C C A G C C T G G G C A A C A C A G C A A 2160
2161 G A C C T T G C T C A A A A C A A A A A A G A C A T A C T A T A A A G T G C T G G C T C C T A T A T A T T 2220
2161 G A C C T T G C T C A A A A C A A A A A A A G A C A T A C T A T A A A G T G C T G G C T C C T A T A T A T T 2220
2221 C A A T A T G T G G T G G A T T C T T G A A T C C T T T C T G A C T C A G A T C T C A T A C G A T T T T T C T G A 2280
2221 C A A T A T G T G G T G G A T T C T T G A A T C C T T T C T G A C T C A G A T C T C A T A C G A T T T T T C T G A 2280
2281 A C T T T T G G A G A A T C C T T G C C T C T C T G C A T T T G C A A A C C G T C A A A G C A C T C C C T T C T G C C 2340
2281 A C T T T T G G A G A A T C C T T G C C T C T C T G C A T T T G C A A A C C G T C A A A G C A C T C C C T T C T G C C 2340
2341 A C C A C A A A G A C A T T T G A T T T T A A A C T T G C A T A T G C T T T G T T C C A A C T T T T A G T A A A 2400
2341 A C C A C A A A G A C A T T T G A T T T T A A A C T T G C A T A T G C T T T G T T C C A A C T T T T A G T A A A 2400
2401 T T A A T C T T G T C A G G G T C T C T G A A C A G C C C T T T A G T C A T A T G C A T T T G A A T A C A T G C C 2460
2401 T T A A T C T T G T C A G G G T C T C T G A A C A G C C C T T T A G T C A T A T G C A T T T G A A T A C A T G C C 2460
2461 C C T A C A G C T 2469
2461 C C T A C A G C T 2469

RESULT 2
US-09-949-016-12284/c
; Sequence 12284, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:

```

; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASES, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12284
; LENGTH: 67156
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-12284

```

Query Match	99.3%	Score 2452.2	DB 4	Length 67156
Best Local Similarity	99.8%	Pred. No. 0		
Matches 2465	Conservative	0	Mismatches	3
			Indels	1
			Gaps	1
QY	1	TTGAGTCTGCTCGAGGCTCCGGGCCAGAGCGGGCGTATTGTTTCACTCGGTGAATGCT	60	
DB	2478	TTGAGTCTGCTTGAGGCTCCGGGCCAGAGCGGGCGTATTGTTTCACTCGGTGAATGCT	2419	
QY	61	CATTTCACGTAAAGAAAACGAGCAACGGCAACGAACTGCGCGAGCGCGCAGACCCCGCA	120	
DB	2418	CATTTCACGTAAAGAAAACGAGCAACGGCAACGAACTGCGCGAGCGCGCAGACCCCGCA	2359	
QY	121	GGGCGCGGTACAGGCACGCTGTGTCCAAACAAGCGCGGAGGCCCGCCCACTCTCC	180	
DB	2358	GGGCGCGGTACAGGCACGCTGTGTCCAAACAAGCGCGGAGGCCCGCCCACTCTCC	2299	
QY	181	CCGACCCGCGCCCGGCCCGCAGCCCTCGCTCGGGGCTCGGACGCAACCGGCACACCT	240	
DB	2298	CCGACCCGCGCCCGGCCCGCAGCCCTCGCTCGGGGCTCGGACGCAACCGGCACACCT	2239	
QY	241	GAGCGAGCGGCGCCCAACCGTAGCGGAGCGCGGTGCGGGAGCGCGCGGGCGGCGTG	300	
DB	2238	GAGCGAGCGGCGCCCAACCGTAGCGGAGCGCGGTGCGGGAGCGCGCGGGCGGCGTG	2179	
QY	301	ACGTACTGTGCGCGCGGGAGCTCAGGGCGGGCGCGGGATGACGGCGCCTCCGCGG	360	
DB	2178	ACGTACTGTGCGCGCGGGAGCTCAGGGCGGGCGGGATGACGGCGCCTCCGCGG	2119	
QY	361	CGAACCGCTGGGACCTCCATTCTGGGGCTGTTTACTCCCAACTCTCGCAGAGCTGGCG	420	
DB	2118	CGAACCGCTGGGACCTCCATTCTGGGGCTGTTTACTCCCAACTCTCGCAGAGCTGGCG	2059	
QY	421	GCCGGGCCAGCGAGGCCCAACAGCTGGGAGCCTCAGCTCCGCCGACCCAGCGTGCCTGTC	480	
DB	2058	GCCGGGCCAGCGAGGCCCAACAGCTGGGAGCCTCAGCTCCGCCGACCCAGCGTGCCTGTC	1999	
QY	481	TGTCCTCGGCTCCCGGGCTTTGCGTGCCTCTGGACCGCGTGGGAGCGGGACACAGC	540	
DB	1998	TGTCCTCGGCTCCCGGGCTTTGCGTGCCTCTGGACCGCGTGGGAGCGGGACACAGC	1939	
QY	541	CGGAGAGTGAACGAAGGTGTCGCGGCAATTGCGCGCGCGGGGCGGTGGCAGGGTGG	600	
DB	1938	CGGAGAGTGAACGAAGGTGTCGCGGCAATTGCGCGCGCGGGGCGGTGGCAGGGTGG	1880	
QY	601	AAGCGAGGGGCGTGGCCAGCAGCTGCCAGCGCGGAGACCGCTGGGGGAAACCTTG	660	
DB	1879	AAGCGAGGGGCGTGGCCAGCAGCTGCCAGCGCGGAGACCGCTGGGGGAAACCTTG	1820	
QY	661	GTCCGCTCTGCGCTCGCTCTAGGATCCCGAAAAGGAGCAGCGGCGCGAAAGCGGCAG	720	
DB	1819	GTCCGCTCTGCGCTCGCTCTAGGATCCCGAAAAGGAGCAGCGGCGCGAAAGCGGCAG	1760	
QY	721	GCTGGGCCAGGATCTAGAAAAGACTGCTCTGGGCGCAGGCTCCCTGTGCCCGCGGGGCTGCTG	780	

1459	CGCAGCGCGCTCGCGGACACCGGTGGTGGGAAACGGAGGACCTTTGTAAACGCCACGCTG	1400
Qy	TTTTGCTCTTTTTGAAAAAACAAGATAAATGTGTAAACTGTCTGAAAAAGCTTGCOCCT	1140
	TTTTGCTCTTTTTGAAAAAACAAGATAAATGTGTAAACTGTCTGAAAAAGCTTGCOCCT	1340
	TTTTGCTCTTTTTGAAAAAACAAGATAAATGTGTAAACTGTCTGAAAAAGCTTGCOCCT	1340
1141	AAAAGATGCTGGGTGACTTAGATGTAGATCAGTTTGTGTTTCAATGTAAATGGACCCAG	1200
Db	AAAAGATGCTGGGTGACTTAGATGTAGATCAGTTTGTGTTTCAATGTAAATGGACCCAG	1280
	AAAAGATGCTGGGTGACTTAGATGTAGATCAGTTTGTGTTTCAATGTAAATGGACCCAG	1280
	AAAAGATGCTGGGTGACTTAGATGTAGATCAGTTTGTGTTTCAATGTAAATGGACCCAG	1280
1201	CCCGGACTCGTACGCGACTAGCAGAGGACTGAAAGCGTCTTCAGGTACTGCTGTGGTGGC	1260
Db	CCCGGACTCGTACGCGACTAGCAGAGGACTGAAAGCGTCTTCAGGTACTGCTGTGGTGGC	1220
	CCCGGACTCGTACGCGACTAGCAGAGGACTGAAAGCGTCTTCAGGTACTGCTGTGGTGGC	1220
	CCCGGACTCGTACGCGACTAGCAGAGGACTGAAAGCGTCTTCAGGTACTGCTGTGGTGGC	1220
1261	GGTGATCGCTACAGGCGGACTCAGACAGTTTTGTGCTTCTGGAACCTGACACTGACCA	1320
Db	GGTGATCGCTACAGGCGGACTCAGACAGTTTTGTGCTTCTGGAACCTGACACTGACCA	1160
	GGTGATCGCTACAGGCGGACTCAGACAGTTTTGTGCTTCTGGAACCTGACACTGACCA	1160
	GGTGATCGCTACAGGCGGACTCAGACAGTTTTGTGCTTCTGGAACCTGACACTGACCA	1160
1321	CGGTAAATGCTGAACCTGCAACCAATATTACAGATCACAGCGCATCATCTTCTTCAACATGA	1380
Db	CGGTAAATGCTGAACCTGCAACCAATATTACAGATCACAGCGCATCATCTTCTTCAACATGA	1100
	CGGTAAATGCTGAACCTGCAACCAATATTACAGATCACAGCGCATCATCTTCTTCAACATGA	1100
	CGGTAAATGCTGAACCTGCAACCAATATTACAGATCACAGCGCATCATCTTCTTCAACATGA	1100
1381	TTTTAAACACAGTTCAGCTTAATATGGTGGATAAATGTAGAAATCAAAATTAACATACCCAC	1440
Db	TTTTAAACACAGTTCAGCTTAATATGGTGGATAAATGTAGAAATCAAAATTAACATACCCAC	1040
	TTTTAAACACAGTTCAGCTTAATATGGTGGATAAATGTAGAAATCAAAATTAACATACCCAC	1040
	TTTTAAACACAGTTCAGCTTAATATGGTGGATAAATGTAGAAATCAAAATTAACATACCCAC	1040
1441	CTCAGGCTTCTACTTCGTAAATTTTGACGAGTTGTTTAACTCTTTGTAGCTCAGCTTCT	1500
Db	CTCAGGCTTCTACTTCGTAAATTTTGACGAGTTGTTTAACTCTTTGTAGCTCAGCTTCT	980
	CTCAGGCTTCTACTTCGTAAATTTTGACGAGTTGTTTAACTCTTTGTAGCTCAGCTTCT	980
	CTCAGGCTTCTACTTCGTAAATTTTGACGAGTTGTTTAACTCTTTGTAGCTCAGCTTCT	980
1501	TCATTAACAAAAATAGGGGTACTAGCCAGGCGGGTGGCTCGCGCTGTAAATCCAGCACT	1560
Db	TCATTAACAAAAATAGGGGTACTAGCCAGGCGGGTGGCTCGCGCTGTAAATCCAGCACT	920
	TCATTAACAAAAATAGGGGTACTAGCCAGGCGGGTGGCTCGCGCTGTAAATCCAGCACT	920
	TCATTAACAAAAATAGGGGTACTAGCCAGGCGGGTGGCTCGCGCTGTAAATCCAGCACT	920
1561	TGGGAGGCGGAGCCGAGCTCACTTGAGGTCAGAAAGTTTCAGACCAAGCTTGTCAAC	1620
Db	TGGGAGGCGGAGCCGAGCTCACTTGAGGTCAGAAAGTTTCAGACCAAGCTTGTCAAC	860
	TGGGAGGCGGAGCCGAGCTCACTTGAGGTCAGAAAGTTTCAGACCAAGCTTGTCAAC	860
	TGGGAGGCGGAGCCGAGCTCACTTGAGGTCAGAAAGTTTCAGACCAAGCTTGTCAAC	860
1621	ATGGGTGAAACGCGGCTCTACCAAAATATAAAACTTAGCTAGTGTGGTAGCGCATG	1680
Db	ATGGGTGAAACGCGGCTCTACCAAAATATAAAACTTAGCTAGTGTGGTAGCGCATG	800
	ATGGGTGAAACGCGGCTCTACCAAAATATAAAACTTAGCTAGTGTGGTAGCGCATG	800
	ATGGGTGAAACGCGGCTCTACCAAAATATAAAACTTAGCTAGTGTGGTAGCGCATG	800
1681	ACTGTAATCCAGCAACTCAGGAGGCTGAGGCAGAGAATCGCTTGAACCTTGGAGGCGGA	1740
Db	ACTGTAATCCAGCAACTCAGGAGGCTGAGGCAGAGAATCGCTTGAACCTTGGAGGCGGA	740
	ACTGTAATCCAGCAACTCAGGAGGCTGAGGCAGAGAATCGCTTGAACCTTGGAGGCGGA	740
	ACTGTAATCCAGCAACTCAGGAGGCTGAGGCAGAGAATCGCTTGAACCTTGGAGGCGGA	740
1741	GGTTGAGTGAGTGAGATCGTACCACTGCACCTCGAGCTTGGGCGACAGAGCGAGCTCT	1800
Db	GGTTGAGTGAGTGAGATCGTACCACTGCACCTCGAGCTTGGGCGACAGAGCGAGCTCT	680
	GGTTGAGTGAGTGAGATCGTACCACTGCACCTCGAGCTTGGGCGACAGAGCGAGCTCT	680
	GGTTGAGTGAGTGAGATCGTACCACTGCACCTCGAGCTTGGGCGACAGAGCGAGCTCT	680
1801	GCCTTAAAAATAAATAAATTTTTTAAAAAAATAGGGGTACTAAATCTTACCTTAAAG	1860
Db	GCCTTAAAAATAAATAAATTTTTTAAAAAAATAGGGGTACTAAATCTTACCTTAAAG	620
	GCCTTAAAAATAAATAAATTTTTTAAAAAAATAGGGGTACTAAATCTTACCTTAAAG	620
	GCCTTAAAAATAAATAAATTTTTTAAAAAAATAGGGGTACTAAATCTTACCTTAAAG	620
1861	GATGAGGTTTAAATTAAGTACACATAAGCCCTTAGCGCAGTGGCTTATGCCCTGTAACT	1920
Db	GATGAGGTTTAAATTAAGTACACATAAGCCCTTAGCGCAGTGGCTTATGCCCTGTAACT	560
	GATGAGGTTTAAATTAAGTACACATAAGCCCTTAGCGCAGTGGCTTATGCCCTGTAACT	560
	GATGAGGTTTAAATTAAGTACACATAAGCCCTTAGCGCAGTGGCTTATGCCCTGTAACT	560
1921	CAACACTTTGGGAGTCTGTGGCGGAGGATCACTTGAAGCCAGAGTTTGAGACTAGTCT	1980
Db	CAACACTTTGGGAGTCTGTGGCGGAGGATCACTTGAAGCCAGAGTTTGAGACTAGTCT	500
	CAACACTTTGGGAGTCTGTGGCGGAGGATCACTTGAAGCCAGAGTTTGAGACTAGTCT	500
	CAACACTTTGGGAGTCTGTGGCGGAGGATCACTTGAAGCCAGAGTTTGAGACTAGTCT	500
1981	GGGCAACAGAGACATGCTCTATAGTTGTGTTTGGTTTGTGTTTAAACAGGTGTGGTGT	2040
Db	GGGCAACAGAGACATGCTCTATAGTTGTGTTTGGTTTGTGTTTAAACAGGTGTGGTGT	440
	GGGCAACAGAGACATGCTCTATAGTTGTGTTTGGTTTGTGTTTAAACAGGTGTGGTGT	440
	GGGCAACAGAGACATGCTCTATAGTTGTGTTTGGTTTGTGTTTAAACAGGTGTGGTGT	440
2041	GTGCACCTGCAGTCCCACTACTAGGAGGCTGAGGTGGAGGACTGCCTGAGCCACGGA	2100
Db	GTGCACCTGCAGTCCCACTACTAGGAGGCTGAGGTGGAGGACTGCCTGAGCCACGGA	380
	GTGCACCTGCAGTCCCACTACTAGGAGGCTGAGGTGGAGGACTGCCTGAGCCACGGA	380
	GTGCACCTGCAGTCCCACTACTAGGAGGCTGAGGTGGAGGACTGCCTGAGCCACGGA	380
2101	GGTCCAGGCTGCAGTGAGCCATGATGTGCCACTGCACTCCAGGCTGGGCAACACAGCAA	2160
Db	GGTCCAGGCTGCAGTGAGCCATGATGTGCCACTGCACTCCAGGCTGGGCAACACAGCAA	320
	GGTCCAGGCTGCAGTGAGCCATGATGTGCCACTGCACTCCAGGCTGGGCAACACAGCAA	320
	GGTCCAGGCTGCAGTGAGCCATGATGTGCCACTGCACTCCAGGCTGGGCAACACAGCAA	320

Qy	2161	GACCTTGTCTCAAAAACAAACAAAGAGCATCTCATAAAGTGGCTCGGCTCCCTATATGATT	2220
Db	319	GACCTTGTCTCAAAAACAAACAAAGAGCATCTCATAAAGTGGCTCGGCTCCCTATATGATT	260
Qy	2221	CAATATGTGGTGGTGAATCTTTGAATCCCTTCCCTGACTCAGATCTCATACGATTTTCTGA	2280
Db	259	CAATATGTGGTGGTGAATCTTTGAATCCCTTCCCTGACTCAGATCTCATACGATTTTCTGA	200
Qy	2281	ACTTTTGGAGATCCCTTGGCTCTCTCGATTTGCAACCGCTCAAGGCACTCCCTCTCTGCC	2340
Db	199	ACTTTTGGAGATCCCTTGGCTCTCTCGATTTGCAACCGCTCAAGGCACTCCCTCTCTGCC	140
Qy	2341	ACCACACAAAGCATTTGATTTTAAACTGACTATGTCCTTCTGTTCCAACTTTAGGTAAA	2400
Db	139	ACCACACAAAGCATTTGATTTTAAACTGACTATGTCCTTCTGTTCCAACTTTAGGTAAA	80
Qy	2401	TTAATCTTGGTCAAGGTTTCTCTGAACAGCCCTTTAGTCTACTATGCCATTGAATACATGGC	2460
Db	79	TTAATCTTGGTCAAGGTTTCTCTGAACAGCCCTTTAGTCTACTATGCCATTGAATACATGGC	20
Qy	2461	CCTACAGCT	2469
Db	19	CCTACAGCT	11

RESULT 4

```

US-09-891-711-1
; Sequence 1, Application US/09891711
; Patent No. 6723553
; GENERAL INFORMATION:
; APPLICANT: Pharmacia AB
; TITLE OF INVENTION: Promoter Sequences
; FILE REFERENCE: 00130
; CURRENT APPLICATION NUMBER: US/09/891,711
; CURRENT FILING DATE: 2001-06-26
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 980
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-891-711-1

```

Query Match

```

Query Match 39.7%; Score 980; DB 4; Length 380;
Best Local Similarity 100.0%; Pred. No. 7.2e-230;
Matches 980; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy	353	CTCGCGCGAAACACACGCTCGGGCACTCCATTTCGGGGCTGTTTACTTCCAACTCTCGCGAG	412
Db	1	CTCGCGCGGAAACACGCTCGGGCACTCCATTTCGGGGCTGTTTACTTCCAACTCTCGCGAG	60
Qy	413	ACTGGGCGGGCGGGCCAGCGAGGCCACACGCTGGGAGCCTCAGCTCCGCGCGACCCACGCT	472
Db	61	ACTGGGCGGGCGGGCCAGCGAGGCCACACGCTGGGAGCCTCAGCTCCGCGCGACCCACGCT	120
Qy	473	GCCCTGTCTGTCGCGCTCTCCGGGGCTTGCGTGGCGCTCTGGACGCCGCTGGGCACGCG	532
Db	121	GCCCTGTCTGTCGCGCTCTCCGGGGCTTGCGTGGCGCTCTGGACGCCGCTGGGCACGCG	180
Qy	533	GACCACGCCGGGAGGATGGACGAAGTGTCTCGGACATTTTCGGCGCGGGGGCGCGGTGG	592
Db	181	GACCACGCCGGGAGGATGGACGAAGTGTCTCGGACATTTTCGGCGCGGGGGCGCGGTGG	240
Qy	593	CAGGGTGGAAACGAGGGGGGTGCCACGCGAGCTGCCAGCGCGCGAGAACGCGCTGGGGG	652
Db	241	CAGGGTGGAAACGAGGGGGGTGCCACGCGAGCTGCCAGCGCGCGAGAACGCGCTGGGGG	300
Qy	653	AACCTTTGGTCCGCTCTGCGCGTCCGCTCTAGGATCCCGAAAAAGGAGCACGGGCGGAAA	712
Db	301	AACCTTTGGTCCGCTCTGCGCGTCCGCTCTAGGATCCCGAAAAAGGAGCACGGGCGGAAA	360
Qy	713	CGCGCCAGGCTGGGCCACGAGATCTTAGAAAGACTGCTCGGCGCAGGCTCCCTGCCCCCGGG	772

361	CGCGCCAGGCTGGGCCAGGATCTAGAAAGACTGCGCTGGCGCAGGCTCCCTGCCCCCGGG	420
773	GCCTGCTGTCAATGGAAGCTCGTGGAGAGCTCGCTTCCGCGCGCGAGACCTTCCTGCAAGGGTC	832
421	GCCTGCTGTCAATGGAAGCTCGTGGAGAGCTCGCTTCCGCGCGAGACCTTCCTGCAAGGGTC	480
833	CAGTCCAGGACACGGCGGCTCGGACACCCACCCCGCGCGCGGACCTCGCCCTGGGTGC	892
481	CAGTCCAGGACACGGCGGCTCGGACACCCACCCCGCGCGGACCTCGCCCTGGGTGC	540
893	CCCTTAACCCGGGCGGTAGCTCGTTAAGATGGCGAAGTGTCCGGTCCGGAAACACGCGAAA	952
541	CCCTTAACCCGGGCGGTAGCTCGTTAAGATGGCGAAGTGTCCGGTCCGGAAACACGCGAAA	600
953	CCCCAAATCCGCGCTGCCCGACCTCTGACCCCGCGCCCCACGGGACGACAGACTGGGGC	1012
601	CCCCAAATCCGCGCTGCCCGACCTCTGACCCCGCGCCCCACGGGACGACAGACTGGGGC	660
1013	TCCCGACGCGCAGCGCGTGC CGGGACACCGGTGCGTGC GGAACCGAGGACCTTTGTTAAC	1072
661	TCCCGACGCGCAGCGCGTGC CGGGACACCGGTGCGTGC GGAACCGAGGACCTTTGTTAAC	720
1073	GCACGTGTTTGCTCTTTTGTGAAAAACACAGAAATAATGTGTATACTGTCTGAAAAAGCT	1132
721	GCACGTGTTTGCTCTTTTGTGAAAAACACAGAAATAATGTGTATACTGTCTGAAAAAGCT	780
1133	TGCGCGCTAAAAGATGCTCGGTGAGCTTAGATGCTAGGATCAGTTTGTGTTTCAATGTAAA	1192
781	TGCGCGCTAAAAGATGCTCGGTGAGCTTAGATGCTAGGATCAGTTTGTGTTTCAATGTAAA	840
1193	TGGACCAAGCCCGGACTCCGTAACGGCACTAGCAGGGGACTGAAAAGCGTCTTCAGGTACTGC	1252
841	TGGACCAAGCCCGGACTCCGTAACGGCACTAGCAGGGGACTGAAAAGCGTCTTCAGGTACTGC	900
1353	TGTTGGGCGGTGATGCGCTACAGCCCGATCAGACAGTTTGTGTCTCTCGGAACCTTGACA	1312
901	TGTTGGGCGGTGATGCGCTACAGCCCGATCAGACAGTTTGTGTCTCTCGGAACCTTGACA	960
1313	CTGCACCAACGGTAATGCTGA	1332
961	CTGCACCAACGGTAATGCTGA	980

RESULTS

```

RESULTS 3
US-09-949-016-13127
; Sequence 13127, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13127
; LENGTH: 32010
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13127

```

Qy	1480	CCTCTTTGTGACCTCAGCTTCTTCAATTACAATAATAGGGGTACTAGCCAGGCGGGTGGCT	1539
Db	6160	CTTCTTGTAGCCTCATCTGCAAAAATGTGGATCAATCACCTTTGGGCCAGGCGCGTGGCT	6219
Qy	1540	CGCGCTGTAAATCCAGCACCTTTGGGAGGCGGAGCAGCCGATCACCTTTGAGGTGAGAAAG	1599
Db	6220	CACACCTGTAAATCCAGCGCTTTTGGAGGCCAAGGCAGGTAGTACCTTTGAGGCCAGGAG	6279
Qy	1600	TTTCAGACCAAGCCTGGTCAAACATCGGTGGAACCGCGGCTCTACCAAAAATATAAAAACTT	1659
Db	6280	TTGGAGACCAAGCCTGGCCAAACAT - GGTGAAACCCCGTCTCTACCAAAAAACAACAAAAATT	6338
Qy	1660	AGCTCAGTGTGGTAGCGATGACTGTAATCCAGCACTCAGGAGGCTGAGGCAG - AGAA	1718
Db	6339	AGCTGGGGTGTGGTGTTTGGTGCTGTAAATCCAGCTACTCGGAGGCTTGAGCAGAGAA	6398
Qy	1719	TCGCTTGAACCTGGAGGCGGAGGTTGAGCTGAGCTGATCGTACCACTGCGACTCCAGC	1778
Db	6399	TTGCTTAAACCTGGGAGGCGAGGTTGAGTAAGCCGAGACCGTGCACATGCAATCCAGC	6458
Qy	1779	TTGGCGACAGAGCGAGACTCTGCCTTAAATAATAATAATAATTTTTTAAAAAAATAGG	1838
Db	6459	CTGGGCAACAGAGCGAGACTCTGCTTAAAAAAAACAAAAAAGGGATCAGGATCA	6518
Qy	1839	GGTACTAATATCTACCTTAAAGGATGAGGGTTAAATTAAGTACACATAAAGCCCTAGCG	1898
Db	6519	CCATGGTCTCTATGCCATGGTTGCTGTGAGGACNTTGGAAATAGCAC - CTGGTGGCCGGCGG	6577
Qy	1899	CAGTGGCTTATGCCTGTAAATCTCAACACTTTTGGGAGTCTGTGGGGGAGGATCACTTGAG	1958
Db	6578	CAGTGTCTCAGCCCTGTAAATCCACAGCACTTTTGGGAGGCTGAGGCGGCGAGATCACTTGAG	6637
Qy	1959	CCCAGGAGTTTGAGACTAGTCTGGGCAACAGAGACATGTCCTATAGTTGTGTTTGGTTT	2018
Db	6638	GCCAGAGTTTGTGAGACCAAGCCTTGGCAACATGGTGAAACCCCGTCTCTACTAAAAATACA	6697
Qy	2019	TGTTTTTACCAGGTGTGGTGTGTCACCTCCAGTCCAGCTACTAGGAGGCTGAGGTG	2078
Db	6698	AAAATTAGCTGGGTGAGGTGGTGCCACCTGTAATCCAGCTACTCGGAGGCTTAGGCA	6757
Qy	2079	GGAGGACTGCTGAGCCAGGAGGTGAGGCTGAGTGCAGTCAGCATGATTTGCCACTGGAC	2138
Db	6758	GGAGAACTACCTTGAACTCGGAGGCGGAGGTTGCAGTGAGCTGAGATGTGCCATTGGAC	6817
Qy	2139	TCAGCCTGGGCAAC - ACAGCAAGACCTTGTCTCAAAAACAAACAAAAAGCAT	2191
Db	6818	TCCAGCTTGGCGCAACACAGCAAACTCTGTCTCAAAAAACAAAAAATAAAAAA	6871

RESULT 6

```

RES001 6
US-09-949-016-12032/c
; Sequence 12032, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12032
; LENGTH: 40855
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12032

```


1 PRIOR FILING DATE: 2000-10-03
1 PRIOR APPLICATION NUMBER: 60/231,498
1 PRIOR FILING DATE: 2000-09-08
1 NUMBER OF SEQ ID NOS: 207012
1 SOFTWARE: FastSeq for Windows Version 4.0
1 SEQ ID NO 17231
1 LENGTH: 187136
1 TYPE: DNA
1 ORGANISM: Human
1 NAME/KEY: misc_feature
1 LOCATION: (1)...(187136)
1 OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17231

Query Match 13.4%; Score 332; DB 4; Length 187136;
Best Local Similarity 70.8%; Pred. No. 1.9e-70;
Matches 487; Conservative 0; Mismatches 190; Indels 11; Gaps 3;
QY 1508 AAAAATAGGGGTACTACCCAGGGGGGTGGCTCGCGCTGTATCCAGCACTTGGGGAG 1567
DB 24550 AAATATTCAGAAAGAGCCAGGCGCGGTGTTTCATCCCTGTAACCCAGCACTTGGGGAG 24491
QY 1568 GCGAGGCGAGCGGATCACTTGAGGTGAGAAAGTTTCAGACCAGCCTGGTCAACATGGGTG 1627
DB 24490 GCTGAGGTGGCGGATGACCTAAGGTGAGCAGTTTGAGACCAGCCTGGCCAACT-GGTG 24432
QY 1628 AAACGGCGGCTCTACCAAAATATAAAACTTTAGCTGAGTGTGGTAGCGCATGACTGTAA 1687
DB 24431 AAACCTGTCTCTACTATAAAATACAAAAGTTAGCTGAGCATGTTGGCTCACATCTGTAA 24372
QY 1688 TCCAGCACTCAGGAGGCTGAGGCA-GAGAACTCGCTTGAACCTGGAGCGGAGGTTGC 1746
DB 24371 TCCAGCTACTCGGAGGGGTGAGGAGGAGAAATCGCTTGAACCCAGGAAGCAGAGGTTGC 24312
QY 1747 AGTGAGCTGAGATCGTACCACTGCATCCAGCTTGGCGGACAGAGCGAGCTCTGCCTTA 1806
DB 24311 AGTGAGCAGAGATCGACCACTTGCACTTCAGCTTGGGCAACAGAGCAAGCTCCGTCTCA 24252
QY 1807 AAAATAAATAAATAAATTTTAAATAAATAAGGGGTACTAATATCTACCTTAAAGGATGAG 1866
DB 24251 AAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 24192
QY 1867 GGTAAATTA-----AGTACACATACAGCTTAGCGGAGTGGCTTATGCCGTGAA 1917
DB 24191 GCAAGCATTTAAATAAATAAATAAATAAAGGCCAGGCACAGTGGCTCATAGCTGTAA 24132
QY 1918 TCTCAACACTTTGGGAGTCTGTGGCGGAGGATCACTTGAGCCAGAGTTTGAGACTAG 1977
DB 24131 TCCAGCACTTCGGGAGGCGGAGGCGAGGCGAGTCACTGAGGTGAGGGTTGAGACCAG 24072
QY 1978 TCTGGGCAACAGAGATGCTCTATAGTTGTGTTTGGTTTGTGTTTGTGTTTGTGTTTGTGTT 2037
DB 24071 CTGGGCAACATGTTGAAACCTGTCTCTACAAAAATAAAAAATTTAGCCAGGCTGGT 24012
QY 2038 GGTGTGCACTGCACTGCTCCAGCTACTAGGAGGCTGAGGTGGAGAGTGCCTGAGCCCA 2097
DB 24011 GGTGCACTGCTATAATCCCAAGCTACTTGGGAGACTGAGGAGAGAAATCACTTGAACCCCG 23952
QY 2098 GGAGGTGCGAGGCTGAGTGAAGCATGATTGTGCACTGCACTCCAGCTGGGCAACACAG 2157
DB 23951 AGAGATGAGAGTTGAGTGAAGTGAATTCGCGCACTGCACTCCAGCTGGCCCAACAGAG 23892
QY 2158 CAAGACTTGTCTCAAAAAACAAACAAA 2185
DB 23891 TGAGACTTGTCAAAAAACAAACAAA 23864

RESULT 9

US-09-949-016-17402/c
; Sequence 17402, Application US/09949016
; Patent No. 681239
; GENERAL INFORMATION:

1 APPLICANT: VENTER, J. Craig et al.
1 TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
1 WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
1 FILE REFERENCE: CL001307
1 CURRENT APPLICATION NUMBER: US/09/949,016
1 CURRENT FILING DATE: 2000-04-14
1 PRIOR APPLICATION NUMBER: 60/241,755
1 PRIOR FILING DATE: 2000-10-20
1 PRIOR APPLICATION NUMBER: 60/237,768
1 PRIOR FILING DATE: 2000-10-03
1 PRIOR APPLICATION NUMBER: 60/231,498
1 PRIOR FILING DATE: 2000-09-08
1 NUMBER OF SEQ ID NOS: 207012
1 SOFTWARE: FastSeq for Windows Version 4.0
1 SEQ ID NO 17402
1 LENGTH: 84425
1 TYPE: DNA
1 ORGANISM: Human
1 US-09-949-016-17402

Query Match 13.4%; Score 330.8; DB 4; Length 84425;
Best Local Similarity 72.3%; Pred. No. 2.6e-70;
Matches 506; Conservative 0; Mismatches 172; Indels 22; Gaps 5;
QY 1494 AGCTTCTTCATTACAAAAATAGGGGTACTAGCCAGCGGGGTGGCTCGCGCTGTAAATCC 1553
DB 39272 AGATTCTGTCTTAAATAAATAAAGATACCGCGCGGCGGTGTTTCATGCTGTAAATCC 39213
QY 1554 CAGCACTTGGGAGGCGGAGGCGGATCACTTTGAGTGCAGAAATTTTTCAGCCAGCCT 1613
DB 39212 CAGCACTTGGGAGGCGGAGGCGGATCAC--GAGTCCAGGAGATTTGAGACCATCT 39155
QY 1614 GGTCAACATGGGTGAAACCCCGCTCTACCAAAATATAAAAACTTTAGCTGAGTGGTA 1673
DB 39154 GGTAAACAT-GGTGAAACCCCGCTCTCTACTAAAAATACAAAAATTTAGCTGGCGGTG 39096
QY 1674 GGCATGACTGTAAATCCAGCAACTCAGGAGGCTGAGGCA-GAGAACTCGCTTGAACCTGG 1732
DB 39095 GAGGTGCTTTGATCTCAGCTACTCAGGAGGCTGAGGAGGAGGAGATGGGTGAAACCCG 39036
QY 1733 GAGGCGGAGGTTTGCAGTGAAGTCCATCCACTGCACTCCAGCTTGGGCGACAGAGC 1792
DB 39035 GAGGCGGAGGTTTGCAGTGAAGTCCATCCACTGCACTCCAGCTTGGGCGACAGAGC 38976
QY 1793 GAGACTCTGCTTAAAAATAAATAAATAAATTTTAAAAATAAATAGGGGTACTAATCTA 1852
DB 38975 AAGACTCCGCTCAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 38932
QY 1853 CCTTAAAGGATGAGGGTTAAATTAAGTACACATAGCCCTAGCGCAGTGGCTTATGCC 1912
DB 38931 ATAAAAAAGATAAAAAACAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 38872
QY 1913 TGTAAATCTCAACACTTTGGGAGTCTGTGGCGGAGGATCACTTTGAGCCCAAGGATTTGAG 1972
DB 38871 TGTAAATCCAGAACTTTGGGAGGCTGATGGGCGCGGATCACTTTAAGTCCAGGAGTTCCAG 38812
QY 1973 ACTAGTCTGGGCAACA--GAGACATGCTCTATAGTTGTGTTTGTGTTTGTGTTTACCAG 2030
DB 38811 ACCAGCTTGGCCAAACATGTTGTAACCTTCTCAATAAATAAATAAATAAATAAATAAATAA 38752
QY 2031 GTGTGGTGTGTGCACTTCAGTCCAGCTACTTAGGGAGGCTGAGGTGGAGGACTGCCT 2090
DB 38751 ATGTGTGTGTGCGACCTGTATCCAGATATCTCAGAGGCTGAGGACCAAGATCGCTC 38692
QY 2091 GAGCCAGGAGGTCGAGGCTGAGGTCAGGTCATGATTTGCCACTGCACTCCAGCCTGGGC 2150
DB 38691 GTACCCAGGAGGCGAGAGGTTGCACTGAGCCAAAGATCTAGCCACTGCACTCCAGCCTGGAT 38632
QY 2151 AACACAGCAAGACCTTGTCTCAAAAAACAAACAAACAAACAAACAAACAAACAAACAAAC 2190
DB 38631 GACAAAGTGAGACTTTTGTCTCAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATA 38592

```

RESULT 10
US-09-949-016-14196
; Sequence 14196, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14196
; LENGTH: 111454
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(111454)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14196

```

Query Match	13.1%	Score 324.2	DB 4	Length 111454	
Best Local Similarity	70.5%	Pred. No. 1.2e-68			
Matches 514	Conservative	0	Mismatches 188	Indels 27	Gaps 5
Qy	1509	AAAAATAGGGGTACTAGCCAGCGGGGTGGCTCGCGCTCTGTAATCCACAGCACTCTGGGGAGG	1568		
Db	37650	AAAAATGGAGACTTGGCGGGCGCAGTGGCTCAGCGCTGTATCGCAGCACTTTGGGAGA	37709		
Qy	1569	CCGAGGCAGCGCGGATCACTTTGAGGTCAGAAGTTTTCAGACCGCTGGTCAAATGGGTGA	1628		
Db	37710	CTGAGCGGGTGGATCACTTTGAGTCAAGAGTTCGAAGACCGCGCTAGCCAAACAT-GGTGA	37768		
Qy	1629	AAGCGCGCTCTACCAAAATATATAAAACCTTACGTCAGTGTGTAGCGCATCAGCTGTAAT	1688		
Db	37769	AACTTTGTCTCTACTAAAAATACAAAATTTAGCTGGGTGTGGTCATGCCCGCTGTAAAT	37828		
Qy	1689	CCCAGCAACTCAGGAGGCTGAGGCA----GAGAAATCGCTTGAACCTGGGAGCGGAGTTT	1744		
Db	37829	CCCAGTTACTCAGGAGGCTCAGGCGCAGGAGGAGAAATGCTTGAACCCAGAAGGTGGAGTTT	37888		
Qy	1745	GCAGTGAAGTCGAGATCGTACCACTGCACCTCCAGCTTTGGGGGACAGAGCAGACTCTGCCT	1804		
Db	37889	TCAGTGAAGCTGAGATCATGCCACTATCTCCAGCCTGGGGCAGAGCAGGAGAAATTTGCCT	37948		
Qy	1805	TAAAAATAAATAAATAATTTTAAAAAAAATAGGGGTACTAATATCTACTCTTAAAGGATG	1864		
Db	37949	AAAAAACAA--AAACAACAATAAAAAAGAGGGGTGGTGGACTTAATCATATAGNAATA	38006		
Qy	1865	AGGGTTAAATTAAGTACACATAAGCCCTAGCGCAGTGGCTTATGCCCTGTAACTCTCAAC	1924		
Db	38007	GTTATTTGC-----AGCCAGGGCGCAGTGGCTTATACCTGTAAATCCAGT	38051		
Qy	1925	ACTTTGGAGTCTGTGGCGGGAGGATCACTTGAAGCCAGGAGTTTCAGACTAGCTCTGGC	1984		
Db	38052	GCITTTTGGAGGCCAAGTTGGAGAGATTCCTTTGAGCCCGAGAGTTCAAGACCAGCTCTGGC	38111		
Qy	1985	AACAGAGACA----TGCTCTATAGTGTGTGTTTGGTTTGTGTTTACCAGGTGTGGTG	2039		
Db	38112	AACATAGCAGACCTCTCTGCACAAAAGTAAAAAAAAGTCAAGCGGTAG	38171		
Qy	2040	TGTGCACCTGCAGTCCAGCTACTAGGGAGGCTGAGTGGGAGGACTGCTGAGCCAGG	2099		
Db	38172	TGTGCGCTGTAGTCCAGCTACTCAGAGGCTGAGTGGGAGGATTTGTTGAGCCACA	38231		

```

Qy 2100 AGGTCAGAGCTCAGTGAGCCATGATTGTGCCACTGCACCTCCAGCTGGGCAACACAGCA 2159
Db 38232 AGGTCAAAGCTCAGTGGAAGCAAGATTGTGCCACTGTACTCAGCTGGGCAAAAGCA 38291
Qy 2160 AGACCTGTGCTCAAAACAAACAAAAGCATACTCATAAAGTGTGCGCTCCTATATGAT 2219
Db 38292 AGACTGTGCTCACAAAAAATAAGGGGGGGTTATTTCGTAAATATATCCTATATA 38351
Qy 2220 TCAATATGT 2228
Db 38352 TTAACATT 38360

RESULT 11
US-09-949-016-12337
; Sequence 12337, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: Fasteq for Windows Version 4.0
; SEQ ID NO 12337
; LENGTH: 111459
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(111459)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12337

```

Query Match	13.1%;	Score 324.2;	DB 4;	Length 111459;
Beat Local Similarity	70.5%;	Pred. No. 1.2e-68;		
Matches 514;	Conservative 0;	Mismatches 188;	Indels 27;	Gaps 5;
QY	1509	AAAATAGGGGTACTAGCCAGGCGGGTGGCTCGCGCTGTATAATCCAGCACACTTGGGGAGG	1568	
DB	37655	AAAATATGGAGACTTGGCGGGCGCAGTGGCTCAGCGCTGTANTCGAGCACCTTTGGGAGA	37714	
QY	1569	CCGAGGCAGCCGGATCACTTTGAGGTGAGAAAGTTTCAGACCGCCTGGTCAACATGGGTGA	1628	
DB	37715	CTGAGGCGGGTGGATCACTTGAGGTCCAGGAGTTCAAGACCGCCTAGCCAAACAT-GGTGA	37773	
QY	1629	AAGCCGGCTCTACCAAAATATAAAACTTAGCTGAGTGTGCTAGCGCATGACACTGTAAT	1688	
DB	37774	AACTTTGTCTCTACTTAAATAATCAAAAANATTAGCTGGGTGTGCTGTGCGCATGTGTAAAT	37833	
QY	1689	CCCAGCAACTCAGGAGGCTGAGGCA-----GAGAATCGCTTGTAACCTGGGAGCGGAGGTT	1744	
DB	37834	CCCAGTTTACTCAGGAGGCTGAGGCAGGAGGAGAAATTGCTTGAAACCCAGAAAGGTGGAGGTT	37893	
QY	1745	GCAGTGAAGCTGAGATCGTACCATCGACCTCCAGCTTGGGCGGACAGCGAGACTCTGGCT	1804	
DB	37894	TCAGTGAAGCTGAGATCATGCGCATATATCTTCAGCCTTGGGCGACAGCGGAGAAATTGGCT	37953	
QY	1805	TAAAAATAAATAATAATTTTTAAAAAAAATAGGGGTACTAATATCTTACCTTTAAAGGATG	1864	
DB	37954	AAAAACAA--AAACAACAATAAAAAAGGGGTGGTGAGCTTAATCATAAAGAAATA	38011	
QY	1865	AGGGTTAAATTAAGTACACATAAGCCCTAGCGCAGTGGCTTTATGCTGTAAATCTCAAC	1924	
DB	38012	GTATATTTC-----AGCCAGGCGCAGTGGCTTTATACCTGTAAATTCAGT	38056	

Db 64222 TACTAAAAACAAAAATAGCGCGGTGGTGGCGCTGTAGTCCAGCTACTT 64163
 QY 1700 AGGAGCTGAGCA-CAGAAATCGTTGAACCTGGGAGCGGAGGTTGCACTGAGCTGAGA 1758
 Db 64162 GGGAGCTGAGGAGGAGAAATGCGTGAACCGGGAGGAGAGCTTGCAGTGGCCGAGA 64103
 QY 1759 TCGTACCACTGCAGCTCGAGCTGGGCGACAGAGCGAGACTCTGCCCTTAAAAATAAATAA 1818
 Db 64102 TCGACCACTGCAGCTCGAGCTGGGCGACAGAGCGAGACTCTCTCTCCAAAAAATAAATAA 64043
 QY 1819 TAATTTTAAAAAATAGGGGTACTTAATATCTTAAAGGATGAGGTTTAAATTAAG 1878
 Db 64042 AAAAAAGAGAGAAAAATGAGAAACAGACTTCTTC-----TGGGATATGTTCAAAA 63991
 QY 1879 TACACACATAAGCCCTAGCGAGTGGCTTATGCTTAAATCTCAACACTTTGGGAGTCTG 1938
 Db 63990 ATGTTATTCAAGGCGGGCAGCGTGGCTCAAGCTTGTAAATCCCAATGCTTTTGAAGTCTG 63931
 QY 1939 TGGCGGGAGGATCACTTGGCCAGGAGTTTGAAGCTAGTCTGGGCAACAGAGACATGTC 1998
 Db 63930 CGGAGGTGATCACTTGGCCAGGAGTTCGTTCAAGACAGGCTAGGCAACATGTTGAACC 63871
 QY 1999 TCTATAGTTGTGTTGTTGTTTATACAGGTGTGGTGTGCACCTGCAGTCCAG 2058
 Db 63870 AGTCTCTACTAAAAATACAAAAATAGCCAGGTGTGGTGTATGCTGTAAATCCAG 63811
 QY 2059 CTACTAGGAGGCTGAGTGGGAGGAGTTCGCTGAGCCAGGAGGTGAGGCTGCAGTGTAG 2118
 Db 63810 CTACTTGGGAGGTGAAATGGGAGGATCGCTTGGCCAGGAGGCGCAAGGTTGCACTGAG 63751
 QY 2119 CCATGATTTGGCCACTGCAGCTCGAGCTGGGCAACAGCAAGACTTGTCTCAAAAAACA 2178
 Db 63750 CCAAGATCCCACTGCAGCTCGAGCTGAGTGACAGATGAGACCTGTCTTTAAAAA 63691
 QY 2179 AACAAAAAGCATACTCATAAAGTGCTC 2205
 Db 63690 AAAAAAATAAAAAAATAAATATTC 63664

RESULT 14
 US-09-949-016-14324/c
 ; Sequence 14324, Application US/09949016
 ; Patent No. 6812339
 ; GENERAL INFORMATION:
 ; APPLICANT: VENTER, J. Craig et al.
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
 ; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
 ; FILE REFERENCE: CL001307
 ; CURRENT APPLICATION NUMBER: US/09/949,016
 ; CURRENT FILING DATE: 2000-04-14
 ; PRIOR APPLICATION NUMBER: 60/241,755
 ; PRIOR FILING DATE: 2000-10-20
 ; PRIOR APPLICATION NUMBER: 60/237,768
 ; PRIOR FILING DATE: 2000-10-03
 ; PRIOR APPLICATION NUMBER: 60/231,498
 ; PRIOR FILING DATE: 2000-09-08
 ; NUMBER OF SEQ ID NOS: 207012
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 14324
 ; LENGTH: 112222
 ; TYPE: DNA
 ; ORGANISM: Human
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)...(112222)
 ; OTHER INFORMATION: n = A,T,C or G
 ; US-09-949-016-14324

Query Match 13.0%; Score 322.2; DB 4; Length 112222;
 Best Local Similarity 70.9%; Pred. No. 3.8e-68;
 Matches 487; Conservative 0; Mismatches 188; Indels 12; Gaps 4;

QY 1520 ACTAGCCAGGCGGGTGGCTCGCGCTGTAAATCCAGCACTTTGGGAGGCCAGGAGCC 1579
 Db 64339 AATGCGGGGCAACGGTGGCTCAAGCTGTAAATCCAGCACTTTGGGAGGCCAGGAGCC 64280
 QY 1580 GGATCACTTGGAGTTCAGAGTTTTCAGACAGCGCTGGTCAACATGGTGAAGCCCGGCTC 1639
 Db 64279 GGATCAC--GAGGTGAGGAGATCGAGACCATCTCGCTAACCA-CAGTGAACCCCGCTC 64223
 QY 1640 TACCAAAAAATATAAAAACTTAGCTGAGTGTGTAGCGCATGACTGTAAATCCAGCAACTC 1699
 Db 64222 TACTAAAAACAAAAAATTAGCCGGCGTGTGGTGGCGCTGTAGTCCAGCTACTT 64163
 QY 1700 AGGAGCTGAGCA-CAGAAATCGTTGAACCTGGGAGGCGGAGGTTGCACTGAGCTGAGA 1758
 Db 64162 GGGAGCTGAGGAGGAGAAATGGCGTGAACCGGGAGGAGAGCTTGCAGTGGCCGAGA 64103
 QY 1759 TCGTACCACTGCAGCTCCAGCTTTGGGCGACAGAGCGAGACTCTGCCCTTAAAAATAAATAA 1818
 Db 64102 TCGACCACTGCAGCTCCAGCTTGGGCGACAGAGCGAGACTCTCTCTCAAAAAAATAAATAA 64043
 QY 1819 TAATTTTAAAAAATAGGGGTACTTAATATCTTAAAGGATGAGGTTTAAATTAAG 1878
 Db 64042 AAAAAAGAGAGAAAAATGAGAAACAGACTTCTTC-----TGGGATATGTTCAAAA 63991
 QY 1879 TACACACATAAGCCCTAGCGAGTGGCTTATGCTGTAAATCTCAACACTTTGGGAGTCTG 1938
 Db 63990 ATGTTATTCAAGGCGGGCAGCGTGGCTCAAGCTTGTAAATCCCAATGCTTTTGAAGTCTG 63931
 QY 1939 TGGCGGGAGGATCACTTGGCCAGGAGTTTGAAGCTAGTCTGGGCAACAGAGACATGTC 1998
 Db 63930 CGGAGGTGATCACTTGGCCAGGAGTTCGTTCAAGACAGGCTAGGCAACATGTTGAACC 63871
 QY 1999 TCTATAGTTGTGTTGTTGTTTATACAGGTGTGGTGTGCACCTGCAGTCCAG 2058
 Db 63870 AGTCTCTACTAAAAATACAAAAATAGCCAGGTGTGGTGTATGCTGTAAATCCAG 63811
 QY 2059 CTACTAGGAGGCTGAGTGGGAGGAGTTCGCTGAGCCAGGAGGTGAGGCTGCAGTGTAG 2118
 Db 63810 CTACTTGGGAGGTGAAATGGGAGGATCGCTTGGCCAGGAGGCGCAAGGTTGCACTGAG 63751
 QY 2119 CCATGATTTGGCCACTGCAGCTCGAGCTGGGCAACAGCAAGACTTGTCTCAAAAAACA 2178
 Db 63750 CCAAGATCCCACTGCAGCTCGAGCTGAGTGACAGATGAGACCTGTCTTTAAAAA 63691
 QY 2179 AACAAAAAGCATACTCATAAAGTGCTC 2205
 Db 63690 AAAAAAATAAAAAAATAAATATTC 63664

RESULT 15

US-09-949-016-17572/c
 ; Sequence 17572, Application US/09949016
 ; Patent No. 6812339
 ; GENERAL INFORMATION:
 ; APPLICANT: VENTER, J. Craig et al.
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
 ; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
 ; FILE REFERENCE: CL001307
 ; CURRENT APPLICATION NUMBER: US/09/949,016
 ; CURRENT FILING DATE: 2000-04-14
 ; PRIOR APPLICATION NUMBER: 60/241,755
 ; PRIOR FILING DATE: 2000-10-20
 ; PRIOR APPLICATION NUMBER: 60/237,768
 ; PRIOR FILING DATE: 2000-10-03
 ; PRIOR APPLICATION NUMBER: 60/231,498
 ; PRIOR FILING DATE: 2000-09-08
 ; NUMBER OF SEQ ID NOS: 207012
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 17572
 ; LENGTH: 113186
 ; TYPE: DNA
 ; ORGANISM: Human
 ; US-09-949-016-17572

Query Match	13.0%;	Score 322.2;	DB 4;	Length 113186;
Best Local Similarity	70.9%;	Pred. No. 3.8e-68;		
Matches 487;	Conservative 0;	Mismatches 188;	Indels 12;	Gaps 4;
QY	1520	ACTAGCAGCGGGTGGCTCGGCCCTGTAATCCAGCACTTGGGGAGGCCGAGGCAGCC	1579	
Db	65326	AATGGCGGGCACGGTGGCTCACGCTGTAATCCAGCACTCTGGGAGGCCGAGCGGGC	65267	
QY	1580	GGATCACTTGAGGTCAAGAGTTTCAGACCGCCTGGTCAACATGGGTGAAACGCCCGGCTC	1639	
Db	65266	GGATCAC--GAGGTCAGGAGATCGAGACCATCTCTGGCTAACAC--CAGTGAACCCCGCTC	65210	
QY	1640	TACCAAAATATAAAACTTAGCTGAGTGTGGTAGCGCATGACTGTAATCCAGCAACTC	1699	
Db	65209	TACTAAAAACAAAAAAATTAGCCGGGCGTGGTGGGGCCCTGTAGTCCCGACTACTT	65150	
QY	1700	AGGAGGCTGAGGCA-GAGAAATCGCTTGAACCTGGGAGGGGAGGTTGCAGTGAGCTGAGA	1758	
Db	65149	GGGAGGCTGAGGCAGGAGAAATGCGCTGAACCGGGAGGCAGAGCTTGCAGTGAGCCGAGA	65090	
QY	1759	TCGTACCACTGCACTCGAGCTTGGGCGACAGAGCGAGACTCTGCCTTAAAAATAAATAA	1818	
Db	65089	TCGACCACTGCACTCGAGCTGGGCGACAGAGCGAGACTCTCTCCAAAAAATAAAAA	65030	
QY	1819	TAATTTTAAAAAATAAGGGTACTAATAATCTACCTTAAAGGATCAGGGTTAAATTAAG	1878	
Db	65029	AAAAAAGAGAGAAAAATGAGAAACAGACTTCTTTC-----TGGGATATGTTCAAAA	64978	
QY	1879	TACACATAAGCCCTAGCGCAGTGGCTTATGCTGTATCTCAACACTTTGGGAGTCTG	1938	
Db	64977	ATGTTATTCAGGCGGGCAGCGTGGCTCAGCGCTGTATCCCAATGCTTTGGAAGTCG	64918	
QY	1939	TGGCGGAGGATCACTTGAGCCCGAGGATTTGAGACTAGTCTGGGCAACAGAGACATGTC	1998	
Db	64917	CGGCAAGTGGATCACTTGAGGCCAGCTGTTCAAGAGCCAGCCTAGGCAACATGGTGAACC	64858	
QY	1999	TCTATAGTTGTGTTTGGTTTGTGTTTACAGGTGGTGGTGTGCACCTGCAGTCCCAG	2058	
Db	64857	AGGTCCTCTATAAAATAAAAAATAGCCAGGTGGTGGTACATGCTGTAAATCCCAG	64798	
QY	2059	CTACTAGGAGGCTGAGGTGGGAGGACTGCTGAGCCCGAGGAGGTGAGGCTGCAGTGAG	2118	
Db	64797	CTACTTGGGAGGCTGAAATGGGAGGATCGCTTGAGCCCGAGGAGCCCAAGGTTGCAGTGAG	64738	
QY	2119	CCATGATTTGCCACTGCAGCTCCAGCTGGGCAACACAGCAAGACCTTGTCTCAAAAAACA	2178	
Db	64737	CCAAGATCCCACCACCTGCAGCTCCAGGCTGAGTGACAGAGTAAGACCCCTGTCTTTAAAAAA	64678	
QY	2179	AACAAAAAGCATACTCATAAAGTCTC 2205		
Db	64677	AAAAAAAAAAAAAAAAAAATATTTC 64651		

Search completed: August 12, 2005, 05:00:21
Job time : 420 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 11, 2005, 21:51:58 ; Search time 1296 Seconds
(without alignments)
11277.655 Million cell updates/sec

Title: US-10-829-118-2

Perfect score: 2469

Sequence: 1 ttgagctgtctggaggtctc.....gaatacatggccctacagct 2469

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq_16Dec04:*

1: Geneseqn1980s:*

2: Geneseqn1990s:*

3: Geneseqn2000s:*

4: Geneseqn2001as:*

5: Geneseqn2001bs:*

6: Geneseqn2002as:*

7: Geneseqn2002bs:*

8: Geneseqn2003as:*

9: Geneseqn2003bs:*

10: Geneseqn2003cs:*

11: Geneseqn2003ds:*

12: Geneseqn2004as:*

13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2469	100.0	2469	6 AAD24181	Aad24181 Human sit
2	980	39.7	980	6 AAD24180	Aad24180 Human sit
3	333.8	13.5	4124	4 AAS01154	Aas01154 5'-untran
4	333.8	13.5	32145	4 AAK68575	Aak68575 Human imm
5	333.8	13.5	32145	4 AAK68491	Aak68491 Human imm
6	332	13.4	110000	13 ABD32780_2	Continuation (3 of
7	330.8	13.4	100445	13 ABD33179	Abd33179 Human can
8	327.4	13.3	147620	10 ADL13739	Adl13739 Osteoarth
9	327.4	13.3	147620	12 ADQ19948	Adq19948 Human sof
10	325.2	13.2	1307	8 ABZ69896	Abz69896 Human end
11	324.8	13.2	110000	12 ADQ97960_2	Continuation (3 of
12	324.2	13.1	110000	11 ACN44150_0	Acn44150 Human gen
13	322.4	13.0	105413	12 ADI36512	Adi36512 Human kin
14	321.6	13.0	91760	11 ACN44410	Acn44410 Human kin
15	321.4	13.0	47219	13 ADR16283	Adr16283 Human Cav
16	321.4	13.0	59884	13 ADR16284	Adr16284 Human Cav
17	319.2	12.9	94719	10 ADE95902	Ade95902 Human STA
18	319.2	12.9	94720	9 ADA02654	Ada02654 Human STA
19	319.2	12.9	94720	10 ADB72392	Adb72392 Human STA
20	318	12.9	24345	13 ADS36503	Ads36503 Human aut

C	21	318	12.9	40947	13	ADS36474	Ads36474 Human aut
C	22	317.8	12.9	192427	10	ADL13825	Adl13825 Osteoarth
C	23	317.2	12.8	270150	11	ADP65796	Adp65796 Human 16p
C	24	316.8	12.8	93390	10	ADD71350	Add71350 Glutamine
C	25	315.8	12.8	25236	4	AAK79879	Aak79879 Human imm
C	26	315.6	12.8	170834	10	AAD62833	Aad62833 Human BAC
C	27	314.4	12.7	109906	6	ABK94411	Abk94411 DNA encod
C	28	314.4	12.7	109906	12	ADL08112	Adl08112 Human gen
C	29	314	12.7	5540	4	AAK79344	Aak79344 Human imm
C	30	313.8	12.7	133893	9	AAD54538	Aad54538 Human pho
C	31	313	12.7	304905	11	ADP75180	Adp75180 Human End
C	32	312	12.6	10770	4	AAK79181	Aak79181 Human imm
C	33	312	12.6	10770	4	AAS32561	Aas32561 Human gen
C	34	312	12.6	12613	4	AAK79178	Aak79178 Human imm
C	35	312	12.6	12613	4	AAS32560	Aas32560 Human gen
C	36	312	12.6	87687	11	ACN45166	Acn45166 Human gen
C	37	311.6	12.6	32249	4	AAL05336	Aal05336 Human rep
C	38	311.6	12.6	32249	4	ABL98205	AbL98205 Human tes
C	39	310.8	12.6	21721	2	AAK83427	Aak83427 Human 1p
C	40	310.8	12.6	39287	6	ABN80533	Abn80533 Human P45
C	41	310.4	12.6	2241	8	ABZ74035	Abz74035 Secrete
C	42	310.4	12.6	2241	8	ADA98642	Ada98642 Human sec
C	43	310.4	12.6	2241	10	ADC20765	Adc20765 Human sec
C	44	310.4	12.6	2241	10	ABZ67622	Abz67622 Human sec
C	45	310.2	12.6	22976	2	AAK83426	Aak83426 Genomic r

ALIGNMENTS

RESULT 1

AAD24181
ID AAD24181 standard; DNA; 2469 BP.
XX
AC AAD24181;
XX
DT 07-MAY-2002 (first entry)
XX
DE Human site-1 protease promoter region.
XX
KW Site-1 protease; S1P; promoter; transcription; obesity; type II diabetes;
KW sterol regulatory element-binding protein; SREBP; hypercholesterolaemia;
KW dyslipidaemia; atherosclerosis; cardiovascular disease; human; ds.
XX
OS Homo sapiens.
XX
PN WO200200873-A1.
XX
PD 03-JAN-2002.
XX
PF 19-JUN-2001; 2001WO-SE001386.
XX
PR 27-JUN-2000; 2000SE-00002417.

(BIOV-) BIOVITRUM AB.

Abrahamson L, Ekblom J, Forsgren M, Hoerling J, Johansson P;

WPI; 2002-139918/18.

Human site-1 protease promoter region for identifying agents capable of inhibiting the promoter activity useful in treating medical conditions such as obesity, diabetes, atherosclerosis and hypercholesterolemia.

Claim 1; Page 15-16; 36pp; English.

The invention relates to human site-1 protease (S1P) promoter region. The promoter sequence is useful for identification of compounds that inhibit transcription of S1P, which in turn results in inhibition of sterol regulatory element-binding protein (SREBP) pathway. The compound identified is useful for the treatment of medical conditions related to obesity, type II diabetes, hypercholesterolemia, dyslipidaemia, atherosclerosis and other cardiovascular diseases. The present sequence

CC	is human SLP promoter region									
XX										
SQ	Sequence 2469 BP; 541 A; 704 C; 719 G; 505 T; 0 U; 0 Other;									
	Query Match 100.0%; Score 2469; DB 6; Length 2469;									
	Best Local Similarity 100.0%; Pred. No. 0;									
	Matches 2469; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	TTGAGTCTGCTGGAGGCTCCGGCCAGAGCAGGCGTATTGTTTCACCTCGGTGAATGCT	60							
DB	1	TTGAGTCTGCTGGAGGCTCCGGCCAGAGCAGGCGTATTGTTTCACCTCGGTGAATGCT	60							
QY	61	CATTTTCAGCTAAAGAAACACAGCAACAGCAACAGCTGCGGAGCGCGAGACCCCGCA	120							
DB	61	CATTTTCAGCTAAAGAAACACAGCAACAGCAACAGCTGCGGAGCGCGAGACCCCGCA	120							
QY	121	GGGCCGGGTACAGGCAACGCTGTGTCCAAACAGCCCGAGAGCCCGCGCCACCTCC	180							
DB	121	GGGCCGGGTACAGGCAACGCTGTGTCCAAACAGCCCGAGAGCCCGCGCCACCTCC	180							
QY	181	CCGACCCGGCCGGCCCGCCGAGCCCTCGCTCGGGCCCTCGGACGCAACCCGACACCT	240							
DB	181	CCGACCCGGCCGGCCCGCCGAGCCCTCGCTCGGGCCCTCGGACGCAACCCGACACCT	240							
QY	241	GAGCGAGCGGCGCCACCGCTAGGCGGAGCGGCTCGGGAGGCGCGCGCGCGCGCTG	300							
DB	241	GAGCGAGCGGCGCCACCGCTAGGCGGAGCGGCTCGGGAGGCGCGCGCGCGCGCTG	300							
QY	301	ACGTACTCTGCGCCCGCGGAGCTCAGGGCGGCGGCGCGGATGACGCGCTCCGCGG	360							
DB	301	ACGTACTCTGCGCCCGCGGAGCTCAGGGCGGCGGCGGCGGATGACGCGCTCCGCGG	360							
QY	361	CGAACAGGCTGGGCACTCAATTCGGGGCTGTTTACTCCCACTCTCGGAGACTGGGG	420							
DB	361	CGAACAGGCTGGGCACTCAATTCGGGGCTGTTTACTCCCACTCTCGGAGACTGGGG	420							
QY	421	GCCGGCCAGCAGAGGCGCCACAGCTGGAGGCTCAGCTCCGCGGACCCAGCTGCGCTG	480							
DB	421	GCCGGCCAGCAGAGGCGCCACAGCTGGAGGCTCAGCTCCGCGGACCCAGCTGCGCTG	480							
QY	481	TGTCGCGCTCCCGGGGCTTGCGTTCGCGCTCTGGACGCGGCTGGGACGCGGACACGC	540							
DB	481	TGTCGCGCTCCCGGGGCTTGCGTTCGCGCTCTGGACGCGGCTGGGACGCGGACACGC	540							
QY	541	CGGAGGATGGACAAAGTGTCTCGACATTTTCGCGCGGCGGGCGCGTGGCAGGGTGG	600							
DB	541	CGGAGGATGGACAAAGTGTCTCGACATTTTCGCGCGGCGGGCGCGTGGCAGGGTGG	600							
QY	601	AAGCGAGGGCGTGGCCAGCAGCTGCCAGGGCGGAGAAACGCGCTGGGGAAACCTTG	660							
DB	601	AAGCGAGGGCGTGGCCAGCAGCTGCCAGGGCGGAGAAACGCGCTGGGGAAACCTTG	660							
QY	661	GTCCGCTCTGCGGCTCGCTTAGGATCCCGAAAGAGCACGCGGCGCGAAAGCGGCCAG	720							
DB	661	GTCCGCTCTGCGGCTCGCTTAGGATCCCGAAAGAGCACGCGGCGCGAAAGCGGCCAG	720							
QY	721	GCTGGCCAGGATCTAGAAAGACTGCTGCGCAGGCTCCCTGCCCGCGGCGCTGCTG	780							
DB	721	GCTGGCCAGGATCTAGAAAGACTGCTGCGCAGGCTCCCTGCCCGCGGCGCTGCTG	780							
QY	781	TCATGAGCTCGTAGAGCTCGCTCCCGCGGACCCCTTCCTGCAAGGGTCCACGCTCA	840							
DB	781	TCATGAGCTCGTAGAGCTCGCTCCCGCGGACCCCTTCCTGCAAGGGTCCACGCTCA	840							
QY	841	GGCACCGGGCTCGGACACCCACCCCGCGCGGCACTGCGCTGGGTGCCCTTTAAC	900							
DB	841	GGCACCGGGCTCGGACACCCACCCCGCGCGGCACTGCGCTGGGTGCCCTTTAAC	900							
QY	901	CCGGCGGTAGCTCGTTAAGATGGGAGAGTGTCCGGTCCGGAAACACCGGAAACCCCAAT	960							
DB	901	CCGGCGGTAGCTCGTTAAGATGGGAGAGTGTCCGGTCCGGAAACACCGGAAACCCCAAT	960							
QY	961	CCCGCTGCCGACCTCCTTGACCCCGGCCCCACGGGACGACAGACTGGGCTCCCGAGC	1020							
DB	961	CCCGCTGCCGACCTCCTTGACCCCGGCCCCACGGGACGACAGACTGGGCTCCCGAGC	1020							

Db 2041 GTGCACCTGCAGTCCAGCTACTAGGAGGCTGAGTGGAGGACTGCCTGAGCCAGGA 2100
 QY 2101 GGTGAGGCTGCAGTGCAGCATGATTTGCGCACTGCACCTCCAGCCCTGGGCAACACAGCAA 2160
 Db 2101 GGTGAGGCTGCAGTGCAGCATGATTTGCGCACTGCACCTCCAGCCCTGGGCAACACAGCAA 2160
 QY 2161 GACCTTGTCTCAAAAAACAAACAAAAGCATCTCATAAAGTGTCTGGCTCCTATATGATT 2220
 Db 2161 GACCTTGTCTCAAAAAACAAACAAAAGCATCTCATAAAGTGTCTGGCTCCTATATGATT 2220
 QY 2221 CAATATGTGGTGGTGAATCTTGAATCTTCTGACTCAGATCTCATAGATCTTCTGA 2280
 Db 2221 CAATATGTGGTGGTGAATCTTGAATCTTCTGACTCAGATCTCATAGATCTTCTGA 2280
 QY 2281 ACTTTTGGAGATCCCTTGGCTCTCTGCAATTTGCAACCGTCAAGGACCTCCCTCTGCC 2340
 Db 2281 ACTTTTGGAGATCCCTTGGCTCTCTGCAATTTGCAACCGTCAAGGACCTCCCTCTGCC 2340
 QY 2341 ACCACACAAAGCATTTGATTTTAAACTTGAATATGCTTCTTCTTCTTCAACTTTAGGTAAA 2400
 Db 2341 ACCACACAAAGCATTTGATTTTAAACTTGAATATGCTTCTTCTTCTTCTTCAACTTTAGGTAAA 2400
 QY 2401 TTAATCTTGGTCAAGGTTCTCTGAAACAGCCCTTTAGTCACTATGCAATGAATATGTC 2460
 Db 2401 TTAATCTTGGTCAAGGTTCTCTGAAACAGCCCTTTAGTCACTATGCAATGAATATGTC 2460
 QY 2461 CCTACAGCT 2469
 Db 2461 CCTACAGCT 2469

RESULT 2

AAD24180
 ID AAD24180 standard; DNA; 980 BP.
 XX AC AAD24180;
 XX DT 07-MAY-2002 (first entry)
 XX DE Human site-1 protease promoter partial sequence.
 XX KW Site-1 protease; S1P; promoter; transcription; obesity; type II diabetes;
 KW sterol regulatory element-binding protein; SREBP; hypercholesterolaemia;
 KW dyslipidaemia; atherosclerosis; cardiovascular disease; human; ds.
 XX OS Homo sapiens.
 XX PN WO200200873-A1.
 XX PD 03-JAN-2002.
 XX PF 19-JUN-2001; 2001WO-SR001386.
 XX PR 27-JUN-2000; 2000SE-00002417.
 XX PA (BIOV-) BIOVITRUM AB.
 XX PI Abrahamson L, Ekblom J, Forsgren M, Hoerling J, Johansson P;
 DR WPI; 2002-139918/18.
 XX PT Human site-1 protease promoter region for identifying agents capable of
 PT inhibiting the promoter activity useful in treating medical conditions
 PT such as obesity, diabetes, atherosclerosis and hypercholesterolemia.
 XX PS Example 1; Page 15; 36pp; English.
 CC The invention relates to human site-1 protease (S1P) promoter region. The
 CC promoter sequence is useful for identification of compounds that inhibit
 CC transcription of S1P, which in turn results in inhibition of sterol
 CC regulatory element-binding protein (SREBP) pathway. The compound
 CC identified is useful for the treatment of medical conditions related to
 CC obesity, type II diabetes, hypercholesterolaemia, dyslipidaemia,

CC atherosclerosis and other cardiovascular diseases. The present sequence
 CC is a fragment of human S1P promoter
 XX Sequence 980 BP; 175 A; 313 C; 326 G; 166 T; 0 U; 0 Other;
 SQ Query Match 39.7%; Score 980; DB 6; Length 980;
 Best Local Similarity 100.0%; Pred. No. 1.5e-207;
 Matches 980; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 353 CTCGCGGCGAACAACGCTTGGGCACTCCATTTGGGGCTGTTACTCCCACTCTCCGAG 412
 Db Seq 1 1 CTCGCGGCGAACAACGCTTGGGCACTCCATTTGGGGCTGTTACTCCCACTCTCCGAG 60
 QY 413 ACTGGGCGCGCGGCGGCGAGGCGCCACAGCTGGGAGCCTCAGCTCCGCGACCCAGCGT 472
 Db 61 ACTGGGCGCGCGGCGGCGAGGCGCCACAGCTGGGAGCCTCAGCTCCGCGACCCAGCGT 120
 QY 473 GGCCTGTGTCTCCCGGCTTCCCGGGCTTTCGCTGGCGCTCTGAGACGCGTGGGCGAGCG 532
 Db 121 GGCCTGTGTCTCCCGGCTTCCCGGGCTTTCGCTGGCGCTCTGAGACGCGTGGGCGAGCG 180
 QY 533 GACCAACCGGCGAGGATGGAACGAGTGTCTCGGACATTTGCGGCGCGGGGCGCGGTGG 592
 Db 181 GACCAACCGGCGAGGATGGAACGAGTGTCTCGGACATTTGCGGCGCGGGGCGCGGTGG 240
 QY 593 CAGGGTGGAAAGCGGAGGCGGCTGGCCAGCGAGCTGCGAGGCGGCGAGAAACGCGCTGGGG 652
 Db 241 CAGGGTGGAAAGCGGAGGCGGCTGGCCAGCGAGCTGCGAGGCGGCGAGAAACGCGCTGGGG 300
 QY 653 AACCTTGTCTCCGCTCTGCGCTCTAGGATCCCGAAAGAGAGACGCGGCGGAA 712
 Db 301 AACCTTGTCTCCGCTCTGCGCTCTAGGATCCCGAAAGAGAGACGCGGCGGAA 360
 QY 713 GGGCGGAGCTGGGCGGAGATCTAGAAAGACTGCGCTGGCGGAGGCTCCCTGCCCGCGG 772
 Db 361 GGGCGGAGCTGGGCGGAGATCTAGAAAGACTGCGCTGGCGGAGGCTCCCTGCCCGCGG 420
 QY 773 GCTGTGTCTATGAGCTCGTGGAGAGCTCGCTTCCCGCGGAGACCTTCTCTGAGGGGTC 832
 Db 421 GCTGTGTCTATGAGCTCGTGGAGAGCTCGCTTCCCGCGGAGACCTTCTCTGAGGGGTC 480
 QY 833 CAGCTCCAGGCAACCGGCGGCTCGGACACCCACCCCGGCGGAGACCTTCCCTGGGTGC 892
 Db 481 CAGCTCCAGGCAACCGGCGGCTCGGACACCCACCCCGGCGGAGACCTTCCCTGGGTGC 540
 QY 893 CCCTTAACCCGGGCGGTAGCTCGTTAAGATGGGAAGTCTCGGTCCGGAACACGCGAA 952
 Db 541 CCCTTAACCCGGGCGGTAGCTCGTTAAGATGGGAAGTCTCGGTCCGGAACACGCGAA 600
 QY 953 CCCCATAATCCCGCTGCGCGACCTCTGACCCCGGCGGCGGACGAGACTTGGGCC 1012
 Db 601 CCCCATAATCCCGCTGCGCGACCTCTGACCCCGGCGGCGGACGAGACTTGGGCC 660
 QY 1013 TCCGACGCGGAGCGGCTGCGGAGACCGGTGCGTGGGAAACGAGGAGCTTTGTAA 1072
 Db 661 TCCGACGCGGAGCGGCTGCGGAGACCGGTGCGTGGGAAACGAGGAGCTTTGTAA 720
 QY 1073 GCCACGTGTTGCTCTTTTGAATAAACAAGATTAATGTTAAACTGTCGAAAGCT 1132
 Db 721 GCCACGTGTTGCTCTTTTGAATAAACAAGATTAATGTTAAACTGTCGAAAGCT 780
 QY 1133 TGCCGCTTAAAGATGCTCTGGGTGACTTAGATGCTTAGGATCAGTTTGTCTTCAATGTAAA 1192
 Db 781 TGCCGCTTAAAGATGCTCTGGGTGACTTAGATGCTTAGGATCAGTTTGTCTTCAATGTAAA 840
 QY 1193 TGGACCAACCGGAGCTCCGTAACGCGGAGCTGAAAGCGCTTTCAGGTACTGC 1252
 Db 841 TGGACCAACCGGAGCTCCGTAACGCGGAGCTGAAAGCGCTTTCAGGTACTGC 900
 QY 1253 TGGTGGGCGGTGATGCGCTACAGCCCGATCAGACAGTTTGTCTTCTTGGAACTTGACA 1312
 Db 901 TGGTGGGCGGTGATGCGCTACAGCCCGATCAGACAGTTTGTCTTCTTGGAACTTGACA 960

667	CGCGAGCTTCAGTGAGCAGAGATCGGCGCACTCCAGCCTGGCGGAAAGAGCGAG	608
2162	ACCTTGTCTCAAAAACAAAACAAAAGCATACTCATAAAGTCTCGGCTCCTATATGATTC	2221
607	ACTCTGTCTCAAAAAAATAAAAAATAAAAAATAAAAACTTATGCTCTTGGCTTACAT	548
2222	AATATGTGGTGGTGATCTTTGAATCCTTTCTGACTCAGATCTCATACGATTTTCTGAA	2281
547	TGTTCTATTGGATGTGCTCTACACCAAAAGTCATTTCCTTTCTGAAACTCTATCACT	488
2282	CTTTTGGAGAA	2292
487	GTTTTGTGTA	477

PR 17-MAR-2000; 2000US-0190076P.
 PR 18-APR-2000; 2000US-0198123P.
 PR 19-MAY-2000; 2000US-0205515P.
 PR 07-JUN-2000; 2000US-0209467P.
 PR 28-JUN-2000; 2000US-02114886P.
 PR 30-JUN-2000; 2000US-0215135P.
 PR 07-JUL-2000; 2000US-0216647P.
 PR 07-JUL-2000; 2000US-0216880P.
 PR 11-JUL-2000; 2000US-0217487P.
 PR 11-JUL-2000; 2000US-0217496P.
 PR 14-JUL-2000; 2000US-0218290P.
 PR 26-JUL-2000; 2000US-0220963P.
 PR 26-JUL-2000; 2000US-0220964P.
 PR 14-AUG-2000; 2000US-0224518P.
 PR 14-AUG-2000; 2000US-0224519P.
 PR 14-AUG-2000; 2000US-0225213P.
 PR 14-AUG-2000; 2000US-0225214P.
 PR 14-AUG-2000; 2000US-0225266P.
 PR 14-AUG-2000; 2000US-0225267P.
 PR 14-AUG-2000; 2000US-0225268P.
 PR 14-AUG-2000; 2000US-0225270P.
 PR 14-AUG-2000; 2000US-0225447P.
 PR 14-AUG-2000; 2000US-0225757P.
 PR 14-AUG-2000; 2000US-0225758P.
 PR 14-AUG-2000; 2000US-0225759P.
 PR 18-AUG-2000; 2000US-0226279P.
 PR 22-AUG-2000; 2000US-0226681P.
 PR 22-AUG-2000; 2000US-0226868P.
 PR 23-AUG-2000; 2000US-0227182P.
 PR 30-AUG-2000; 2000US-0228924P.
 PR 01-SEP-2000; 2000US-0229287P.
 PR 01-SEP-2000; 2000US-0229343P.
 PR 01-SEP-2000; 2000US-0229344P.
 PR 01-SEP-2000; 2000US-0229345P.
 PR 05-SEP-2000; 2000US-0229509P.
 PR 05-SEP-2000; 2000US-0229513P.
 PR 06-SEP-2000; 2000US-0230437P.
 PR 06-SEP-2000; 2000US-0230438P.
 PR 08-SEP-2000; 2000US-0231242P.
 PR 08-SEP-2000; 2000US-0231243P.
 PR 08-SEP-2000; 2000US-0231244P.
 PR 08-SEP-2000; 2000US-0231413P.
 PR 08-SEP-2000; 2000US-0231414P.
 PR 08-SEP-2000; 2000US-0232080P.
 PR 08-SEP-2000; 2000US-0232081P.
 PR 12-SEP-2000; 2000US-0231968P.
 PR 14-SEP-2000; 2000US-0232397P.
 PR 14-SEP-2000; 2000US-0232398P.
 PR 14-SEP-2000; 2000US-0232399P.
 PR 14-SEP-2000; 2000US-0232400P.
 PR 14-SEP-2000; 2000US-0232401P.
 PR 14-SEP-2000; 2000US-0233063P.
 PR 14-SEP-2000; 2000US-0233064P.
 PR 14-SEP-2000; 2000US-0233065P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 25-SEP-2000; 2000US-0234998P.
 PR 26-SEP-2000; 2000US-0235484P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 27-SEP-2000; 2000US-0235836P.
 PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239935P.

PR 13-OCT-2000; 2000US-0239937P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 20-OCT-2000; 2000US-0241826P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246603P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249219P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249264P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-483426/52.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides, useful for preventing, diagnosing and/or treating cancers and metastasis.

Disclosure; SEQ ID NO 23303; 3071pp + Sequence Listing; English.

AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I) amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic activity, and can be used in gene therapy and vaccine production. (I)

CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patient's own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting the
 CC nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/hematopoietic-related diseases, especially
 CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
 CC to AAK87694 represent human immune/hematopoietic antigen genomic
 CC sequences from the present invention. AAK54950 and AAK82169
 CC represent sequences used in the exemplification of the present invention
 XX
 SQ Sequence 32145 BP; 8777 A; 7496 C; 7117 G; 8755 T; 0 U; 0 Other;

Query Match		13.5%;	Score 333.8;	DB 4;	Length 32145;
Best Local Similarity		66.4%;	Pred. No. 1.1e-63;		
Matches		525;	Conservative	0;	Mismatches 262; Indels 4; Gaps 3;
QY	1503	ATTACAAAATAGGGTACTAGCAGCGGGGGTGGCTGCGGCTGTAAATCCGACACTTG	1562		
DB	3586	ATTGAAAAATAAATTTCAAGGCGAGGACAGTGGCTCAGCGCTGTAAATGCTAGCACTTT	3527		
QY	1563	GGGAGCGGAGCGGAGCGGATCACTTGGTCAAGAGTTTCAGACAGCGCTGGTCAACAT	1622		
DB	3526	GGGAGCGGAGCGGAGCGGATCACTTGGTCAAGAGTTTCAGACAGCGCTGGTCAACAT	3467		
QY	1623	GGGTGAAACCGCGGCTCTACCAAAAATATAAAAATTTAGCTGAGTGGTAGCGCATGAC	1682		
DB	3466	TAGTGAAACCGCTGCTCTACTAAAAACAGAAA--TTAGCTGGGCATGGTGGCAGCACC	3408		
QY	1683	TGTAATCCGACAACTCAGAGGCTGAGGCA--GAGAATCGTTTGAACCTGGGAGCGGAG	1741		
DB	3407	TGTAATCCGACGACTCTGGAGGCTGAGGCGAGGAATCGTTGAACCTGGGAGGTGGAG	3348		
QY	1742	GTTCAGTGGAGTGTAGATCGTACCACTGCACCTCCAGCTTGGGCGAGAGCGAGACTCTG	1801		
DB	3347	GTTCAGTGGAGTGTAGATCGTACCACTGCACCTCCAGCTTGGGCGAGAGCGAGACTCTG	3288		
QY	1802	CCTTAAAAATAAATAAATTTTAAAAAAAATAGGGGTACTTAATCTACCTTAAAGG	1861		
DB	3287	TCTCAAAATAAATAAATAAATAAATAAATTTTCTTCTTCTTCTTCTTCTTCTTCTTCTT	3228		
QY	1862	ATGAGGGTTAAATTAAGTACACATAGCCCTAGCGAGTGGCTTATGCTGTATCTC	1921		
DB	3227	ATGATGTTTCTAGAAAATAAAAC--TAGCGCGGGGTGGTGGCTCAGCGCTGTAATCCC	3170		
QY	1922	AACACTTTGGGAGTCTGTGGCGGAGGATCACTTGGAGCCAGAGTTTGAGACTAGTCTG	1981		
DB	3169	AGCACTTTGGGAGGCGGAGGAGCGGATCACAGAGTTCAGGAGTCAAGACCATCTCTGGC	3110		
QY	1982	GGCAACAGAGACATGCTCTATAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG	2041		
DB	3109	TAAACAGGTGAACCCCGTCTCTACTGAAAATAAATAAATAAATAAATAAATAAATAAATAA	3050		
QY	2042	TGCACCTGCACTGCCAGTCTAGGAGGCTGAGGTGGGAGGACTGCCTGAGCCGAGGAG	2101		
DB	3049	GGCACCTGTAATCCAGCTACTTTGGAGGCTGTAGGCGAGGAGATGGCGTGAACCCCGGAG	2990		
QY	2102	GTGAGGCTGCACTGAGCCATGATTTGGCCACTGCACTCCAGCTGGGCGAACAACAGCAAG	2161		
DB	2989	GCGAGGCTTGCAGTGAAGAGATTCGCGCCACTGCACTCCAGCTGGGCGAAGAGCGGAG	2930		
QY	2162	ACCTGTCTCAAAAACAAACAAAGCATACTCATAAAGTGTCTCGGCTCTATATGATTC	2221		
DB	2929	ACTCTGTCTCAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	2870		
QY	2222	AATATGTGGTGGTATCTTGAATCCTTTCTGACTCAGATCTCATACGATTTCTGAA	2281		
DB	2869	TGTTTCTATTGGATGTGCTCTACACCAACCGTCAATTCCTCTTCTGAAACTCTATCACT	2810		

QY	2282	CTTTTGGAGAA	2292		
DB	2809	GTTTTGTATA	2799		
RESULT 6					
ABD32780_2/c					
Continuation (3 of 5) of ABD32780 from base 200001 (Human cancer-associated genomic DNA)					
WP Sequence split into 5 fragments LOCUS ABD32780 Accession ABD32780					
WP	Fragment Name	Begin	End		
WP	ABD32780_0	1	110000		
WP	ABD32780_1	100001	210000		
WP	ABD32780_2	200001	310000		
WP	ABD32780_3	300001	410000		
WP	ABD32780_4	400001	430442		
Query Match					
Best Local Similarity 13.4%; Score 332; DB 13; Length 110000;					
Matches 487; Conservative 0; Mismatches 190; Indels 11; Gaps 3;					
QY	1508	AAAAATAGGGTACTAGCCAGCGGGGTGGCTCGCGCTGTAAATCCGACACTTGGGGAG	1567		
DB	59424	AAATATTCAGAAAGAGGCGAGCGCGGTGTTTCATGCTGTAAACCCGACACTTTGGGAG	59365		
QY	1568	GCGAGGCGAGCGGATCACTTGGGTCAAGAGTTTCAGACACGCTGTGTCAACATGGGTG	1627		
DB	59364	GCTGAGTGGGCGGATGACCTAAGGTGAGCAGTTTGAGACGAGCTTGGCCACAT--GGTG	59306		
QY	1628	AAACCGCGCTCTACCAAAAATATAAAATTTAGCTGAGTGGTGAAGCATGACTGTAA	1687		
DB	59305	AAACCGCTCTCTACTTAAAAATACAAAAAGTTAGCTGAGCATGGTGGCTCACAATCTGTA	59246		
QY	1688	TCCGAGCACTCAGGAGGCTGAGCA--GAGATCGCTTGAACCTGGAGCGGAGGTTCG	1746		
DB	59245	TCCGAGCTACTCGGAGGCTGAGGAGGAGATCGCTTGAACCCGAGCAGAGGTTCG	59186		
QY	1747	AGTGAGCTGAGATCGTACCACTGCACCTCGAGCTTGGGCGACAGAGCAGACTCTGCCTTA	1806		
DB	59185	AGTGAGCAGAGATCGACCACTTGCATTCAGCTTGGGCAACAGAGCAAGACTCCGTCCTCA	59126		
QY	1807	AAATAAATAAATAAATTTTAAAAAAAATAGGGGTACTTAATATCTTAAAGGATGAG	1866		
DB	59125	AAAACAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA	59066		
QY	1867	GGTAAATTA-----AGTACACACATAAGCCCTAGCGAGTGGCTTATGCCCTGTA	1917		
DB	59065	GCAAGCATTTAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	59006		
QY	1918	TCTCAACACTTTGGGAGTCTGTGGCGGAGGATCACTTGGCCGAGGAGTTTGAGACTAG	1977		
DB	59005	TCCAGCACTTCGGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	58946		
QY	1978	TCTGGGCAACAGAGACATGCTCTATAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	2037		
DB	58945	CCTGGCCAACTGGTGAACCTGTCTACAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAA	58886		
QY	2038	GGTGTGCACTGAGTCCGAGCTACTAGGGAGGCTGAGTGGGAGGAGTGCCTGAGCCCA	2097		
DB	58885	GGTGTGCTGTATAATATCCAGCTACTTGGGAGAGTGGGAGAGAGAGAGAGAGAGAGAGAGAG	58826		
QY	2098	GGAGGTGCGAGGCTGCACTGAGGAGGATGATTTGTGCACTGCACTCCAGCTGGGCAACACAG	2157		
DB	58825	AGAGATGAAGGTTTGCAGTGAAGTGGCCACTGCACTCCAGCTTGGCCCTGGCCCAACAGAG	58766		
QY	2158	CAAGACCTTGTCTCAAAAAACAAAAA	2185		
DB	58765	TGAGACCTTGTCAAAAAAATAAAAAA	58738		
RESULT 7					
ABD33179/c					
ID	ABD33179	standard; DNA; 100445 BP.			
XX					

CC associated with a disease, preferably osteoarthritis. The cell line and
CC the non-human animal are useful for screening for an agent for diagnosing
CC an individual having susceptibility to joint space narrowing and/or
CC osteophyte development and/or joint pain. This sequence corresponds to
CC the polynucleotide encoding a protein listed in the specification. (Note:
CC The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format directly from WIPO at
CC [ftp.wipo.int/pub/published_pct_sequences](http://www.int/pub/published_pct_sequences)).

XX
SQ Sequence 147620 BP: 39001 A: 34978 C: 35560 G: 38081 T: 0 U: 0 Other:

Sequence 147620 BP; 39001 A; 34978 C; 35560 G; 38081 T; 0 U; 0 Other;

Query Match	13.3%	Score 327.4;	DB 10;	Length 147620;
Best Local Similarity	71.0%;	Pred. No. 4.6e-62;		
Matches 489;	Conservative	0;	Mismatches 196;	Indels 4;
			Gaps	4;

Qy	1504	TTACAAAAATAGGGGTACTAGCCAGCGGGGTGGCTCGCGCTGTAATCCACGACCTTGG	1563
Db	4080	TCAAAAAATAATAAAAATAGCGCGGTGTGGTGGCTCACGCTGTAATCTCAGCACCTTGG	40139

Qy	1564	GGAGGCCGAGCGACCGGATCACTTGAGGTGAGAGTTTCAGACGAGCCTGGTCAACATG	1623
Db	40140	GGAGGCCGAGCGGGCGGATCACTTGAGGTGAGAGTTCAAGACGAGCCTGGCCAAAT-	40198

Qy	1624	GGTGAACCGCGGCTCTACCAAAAATATAAAAACCTTAGCTGAGTGTGGTAGCGCATGACT	1683
Db	40199	GGTGAACCGCGTCTCTACTAAAAATACAAAAA-TTAGCTGGGCGCTGGTGGTGGTGCCT	40257

Qy 1684 GTAATCCAGCAACTCAGGAGCTGAGGCA - GAGAAATCGCTTGAACTCTGGAGGCGGAGG 1742

Db 40258 CTAATCCAGTACTCTGGAGGCTGAGGCAAGAAATCACTTGAATCCAGGAGGCGGAGG 40317

Qy	1743	TTGCAGTGAGCTGAGATCGTACCACTGCCTCCAGCTTGCGCGAGCGAGACTCTGC	1802
Dβ	40318	TTGCAGTGAGCTAGATCATGCCACTGCATTCACGCTGGTGATAGAGCGAGACTCAGT	40377

[illegible]

QY
DB

1863 TGAGGGTTAAATTAGTACACACATGAAGCCCTAGGCAGTGGCTTATGCGTGTAATCTCA 1922

40438 AGTAGAAAAAACAATAATGACGACGACACATCACTGTCTGGCGCTATGGCTGTATTTCTC 40496

[illegible]

QY	1983	GCAACAGAGACATGCTCTATAGTTCTGTTGGTTTGTGTTTACGAGGTGTGGTGT	2042
1983			
1984			
1985			
1986			
1987			
1988			
1989			
1990			
1991			
1992			
1993			
1994			
1995			
1996			
1997			
1998			
1999			
2000			
2001			
2002			
2003			
2004			
2005			
2006			
2007			
2008			
2009			
2010			
2011			
2012			
2013			
2014			
2015			
2016			
2017			
2018			
2019			
2020			
2021			
2022			
2023			
2024			
2025			
2026			
2027			
2028			
2029			
2030			
2031			
2032			
2033			
2034			
2035			
2036			
2037			
2038			
2039			
2040			
2041			
2042			

QY 2043 GCACCTGCAGTCCCAGCTACTAGGGAGGCTGAGGTGGGAGGACTCCCTGAGCCCGAGG 2102

Qy 2103 TCGAGGCTGCAGTGAACCATGATTGTGCCACTGCACTCGACCTCGGGCACACAGCAAG 2162

Qy 2163 CCTGTCTCAAAAACAAAAGCATA 2191

RESULT 9
ADQ19948/c
ID ADQ19948 standard; DNA; 147620 BP.

XX soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
KW ds.

Query Match	13.3%	Score	327.4	DB	12	Length	147620
Best Local Similarity	71.0%	Pred. No.	4.6e-62				
Matches	489	Conservative	0	Mismatches	196	Indels	4
						Gaps	4

Qy	1504	TTACAAAAATAGGGTACTAGCCAGCGCGGGTGGCTCGCGCCCTGTAAATCCCGACACTTGG	1563
Db	107541	TCAAAAAATAATAAAAATAGGCGGGTGGTGGCTCAGCCCTGTAAATCTCAGACATTTG	107482

Qy 1564 GGAGGCCGAGGCGCGGATCACTTGAGGTCAAGATTTTCAGACCAGCCTCGTGGTCAACATG 1623

Db 107481 GGAGGCCGAGGCGGCGCGATCACTTGAGGTCAAGATTTTCAGACCAGCCTCGTGGCACAACAT- 1074223

Qy	1624	GGTGAACCGCGGCTCTACCAAAATATATAAACTTAGCTAGTGTGGTGGCGCATGACT	1683
Dp	107422	GGTGAACCGCGTCTCTACTAAAATACAAAAA-TTAGCTGGGCGTGGTGGGTGGCT	107364

QY	1684	GTAATCCGAGCAACTCAGGAGGTGAGCCA-GAGAATCGCTTGAACTCTGGGAGCGCGAGG	1742
DP	107363	CTAATCCGAGCTACTTGGGAGGTGAGCACAAGAATCACTTGAATCCAGGAGCGCGAGG	107304

QY	1743	TTGCAGTGAGCTGAGATCGTACCACTGCATCCAGCTTTGGGCGACAGAGCGAGACTCTGC	1802
DB	107303	TTTCAGTGAGGCTTGAATCATGCCACTGCATTCAGCCCTGGGTGATAGAGCGAGACTCAGT	107244

[illegible]

QY 1863 TGAGGGTTAAATTTAAGGTACACACATAGCCCTAGCGAGTGGCTTATGCCTGTAATCTCA 1922

Db 107183 AGAAGAAAAAGAAAGAGA-AGAAACAGCCAGGTGTCAGTGGCTCATGCTGTAAATCCA 107125
QY 1923 ACACCTTTGGAGTCTGTGCGGAGGATCACTTTGAGCCAGAGTTTGAGACTAGTCTGG 1982
Db 107124 ACACCTTTGGAGGCTGAGATGGTGGATCACTTTGAGTCAAGATTCGAGACAGCCTGG 107065
QY 1983 GCAACAGAGACATGCTCTATATAGTTGTGTTTGGTTTGTATTTTACAGGTGGTGGTGT 2042
Db 107064 CCAACATGTTGAACCCCGTCTCTACTAAAAATACCAAAATCAGCCAGACATGTTAGCAT 107005
QY 2043 GCACCTGACGTCCAGCTACTAGGAGGCTGAGGTGGAGGACTGCTGAGCCAGGAGG 2102
Db 107004 GCATCTGTAATCCAGCTACTTTGGGAGGCTGATGCAGGAGAAATCACTTGAACCTTGGGAGG 106945
QY 2103 TCGAGGCTCCAGTACCATGATTTGGCCACTGCACCTCCAGCTGGGCAACACAGCAAGA 2162
Db 106944 CGGAGTTTGCAGTGAAGCCGAGATCACGCCATTGTACTCCAGCCTGGCAACGAGGAGCA 106885
QY 2163 CTTTGTCTCAAAAAACAAACAAAAAGCATA 2191
Db 106884 CTCATAAACAAAAAAGGAAAAAGAAA 106856

RESULT 10
ABZ69896/c
ID ABZ69896 standard; cDNA; 1307 BP.
XX AC ABZ69896;
XX DT 23-APR-2003 (first entry)
XX DE Human endothelial cell differentiation factor protein 9.35 cDNA.
XX KW Human; endothelial cell differentiation factor protein 9.35; EDF-1;
XX KW cancer; HIV; gene; ss.
XX OS Homo sapiens.
XX FT Key Location/Qualifiers
XX CDS 351..608
XX FT /*tag= a
XX FT /product= "EDF-1 9.35"
XX CN1363571-A.
XX PD 14-AUG-2002.
XX 05-JAN-2001; 2001CN-00105043.
XX 05-JAN-2001; 2001CN-00105043.
XX (BODE-) BODE GENE DEV CO LTD SHANGHAI.
XX Mao Y, Xie Y;
XX WPI; 2003-000312/01.
XX P-PSDB; ABP1850.
XX Polypeptide-human endothelial cell differentiation factor EDF-1 protein
XX 9.35 and polynucleotide for coding it.
XX Claim 6; Page 25-26 (Disclosure); 32pp; Chinese.
XX The invention relates to the novel human endothelial cell differentiation
XX factor (EDF-1) protein 9.35, and the polynucleotide encoding it. The
XX polypeptide is useful in treating diseases such as cancer and HIV
XX infection. The antagonist of the polypeptide and its medical function,
XX and the application of the polynucleotide are also disclosed. The present
XX sequence encodes the human EDF-1 protein 9.35 of the invention

XX SQ Sequence 1307 BP; 285 A; 327 C; 323 G; 372 T; 0 U; 0 Other;
Query Match 13.2%; Score 325.2; DB 8; Length 1307;

Best Local Similarity 71.0%; Pred. No. 3.7e-62;
Matches 487; Conservative 0; Mismatches 193; Indels 6; Gaps 4;
QY 1506 ACAAATAATAGGGTACTAGCCAGCGGGTGGCTCGCCCTGTATATCCAGCAGCTTGGG 1565
Db 1070 AAATAATTAATAAACAGCCCAAGCGGTGGTTCACGCTGTATATCCAGCAGCTTGGG 1011
QY 1566 AGCGCAGCGCAGCCGATCACTTTGAGTTCAGAAAGTTTCAGACAGCCTGGTCAACATGGG 1625
Db 1010 AGCCCAAGCGCGGAGATCACTGAGGTTCAGTAGTTCAGAGACCCCTAGCTTAACAT-GG 952
QY 1626 TGAACCGCGGCTCTACCAAAATATAAAATTAAGCTGAGTTCAGTTCAGCAGCTGCT 1685
Db 951 TGAACCCCGCTCTCAACTAAAAAATACAAAAAATGATCGGGTGTGGCGCGCGCTGT 892
QY 1686 AATCCCAACCACTCAGGAGGCTGAGGCA-GAGAAATCGCTTGAACCTGGGAGCGGAGGTT 1744
Db 891 AATCCCACTACACGGAGCTGAGGAGGAGATCGCTTAAGCCCGGAGGAGCGAGGTT 832
QY 1745 GCAGTGAGCTGAGATCGTACCACTGCACCTCCAGCTTGGGCGACAGAGGAGACTCTGCCT 1804
Db 831 GCAGTGAGCTGAGATCGCGCCACTGCACCTCCAGCCTGGCGACAAAAAGCAGACTCTGTCT 772
QY 1805 TAAAAATAAATAAATATTTTAAAAAATAGGGTACTAATATATCTACCTTAAAGGATG 1864
Db 771 TAAAAAATAAATATATATATATATATATATATATCTCTACAAAAAATTCCTAGATCTC 712
QY 1865 AGGGTTAAATTAAGTACACACATAAGCCCTAGCCAGCTGGCTTATGCCCTGTATCTCAAC 1924
Db 711 TGACCACAAAAATCCGCCCCCTAGCTGGGCGGGTGGCTCAGCCCTGTATATCCAGC 652
QY 1925 ACTTTGGGAGTCTGTGGCGGAGGATCACTTGAGCCCAAGGAGTTTGAGACTAGTCTGGGC 1984
Db 651 ACTTTGGGAGGCGGAGGTGGTGGATCACTCAGGTTCAGAGGTTCAAGACCAAGCCTGGCC 592
QY 1985 AACAGAGACATGCTCTATAGTTGTGTTTGGTTTGTGTTTGA-CCAGTGTGGTGGTGG 2043
Db 591 AACATGGTGAACCCCTCTCTACTAAAAAATACAAAAATCAGCCGGGTGTGGTGGCGGG 532
QY 2044 CACCTGCAGTCCAGCTACTAGGAGGCTGAGTGGGAGGCTGCTGAGGCC---AGGA 2100
Db 531 CGCCTGTAAATCCAGCTACACGGAGGCTGAGGAGGAGAAATCGCTTGAACCCGGGAGGA 472
QY 2101 GGTGAGGCTGCAGTGAGCCATGATTTGTGCCACTGCACCTCCAGCCTGGGCAACACAGCAA 2160
Db 471 GACAGAGGTTGAGTGAGCCAAAGATCGGCCATTGCACTCCAGCCTGGGCGCAGAGCGA 412
QY 2161 GACCTTGTCTCAAAAAACAAACAAAAA 2186
Db 411 AACACCGCTCTTAAAAAATAAAAAA 386

RESULT 11
ADQ97960.2/c
Continuation (3 of 4) of ADQ97960 from base 200001 (Human cancer associated sequence HDL)
WP Sequence split into 4 fragments LOCUS ADQ97960 Accession Adq97960

WP	Fragment Name	Begin	End
WP	ADQ97960_0	110000	110000
WP	ADQ97960_1	100001	210000
WP	ADQ97960_2	200001	310000
WP	ADQ97960_3	300001	390183

Query Match 13.2%; Score 324.8; DB 12; Length 110000;
Best Local Similarity 68.6%; Pred. No. 1.6e-61;
Matches 528; Conservative 0; Mismatches 222; Indels 20; Gaps 5;
QY 1524 GCAGCGCGGGTGGCTCGCCCTGTATATCCAGCAGCTTGGGAGGCGGAGCGAGCGGAT 1583
Db 15258 GCCAGCGCGGTGGCTCAGCCCTGTATATCCAGCAGCTTTGGGAGGCTGAGGCGGGGAT 15199
QY 1584 CACTTGAGGTCAAGATTTTCAGACAGCCTGTCACATCGGTGAAACCGCGCTTACC 1643
Db 15198 CA--TAAGGTTCAGGAGATCAAGACCGCCTGGGTAACA-CGATGAAATCCATCTCTACT 15142

DR WPI; 2003-328604/31.
XX Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT comprises a nucleotide sequence.
XX
XX
XX Claim 1; SEQ ID NO 844; Opp; English.
XX
XX The present invention relates to novel DNA and protein sequences which
CC are associated with carcinomas. The sequences are useful for: (i) for
CC screening drug candidates; (ii) for screening of bioactive agent capable
CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC a bioactive agent capable of modulating the activity of CAP; (iv) for
CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC determining Carcinoma Associated (CA) gene copy number. In addition, the
CC CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC carcinoma including lymphoma. The present sequence is one such CA coding
CC sequence. Note: This patent is an equivalent to basic patent
CC US2002182586A1, for which no sequence data was published
XX
SQ Sequence 91760 BP; 21177 A; 23589 C; 24129 G; 22845 T; 0 U; 20 Other;
Query Match 13.0%; Score 321.6; DB 11; Length 91760;
Best Local Similarity 72.1%; Pred. No. 7.7e-61;
Matches 490; Conservative 0; Mismatches 179; Indels 11; Gaps 5;
QY 1523 AGCCAGCGGGGCTGCGCGCTGTAATCCAGCACTTGGGAGCGCGAGCGCGGA 1582
DB 68876 AGCGGGCGCAGTGGCTCAGCGCTGTAATCCAGCACTTGGGAGCGCGAGCGAGTGA 68817
QY 1583 TCACCTGAGTCAAGATTTCAGACCAGCTGTCAACATGGGTGAACCGCGCTCTAC 1642
DB 68816 TCAC--GAGGTCAAGCTTTTGAGACCAGCGCTGTCAACAT--GGTGAACCGCGCTCTAC 68760
QY 1643 CAAAAATATAAAACTTAGCTGAGTGTGGTAGCGCATGCTGTAATCCAGCAACTCAGG 1702
DB 68759 TAAATATACAAATATTAGCCAG--GCACGGTGGCAGCGCTTAACCCCACTACTCCGG 68703
QY 1703 AGGCTGAGGC--AGAGAAATCGCTTGAACCTCGGAGGCGGAGGTGCGAGTGGAGATCG 1761
DB 68702 AGGCTGATGCGGAGAACTACTTTGAACCGAGGAGGTGAAGGTTGCGAGTGAACCAAGACCG 68643
QY 1762 TACCACTGCACTCCAGCTTGGGCGACAGAGCGAGCTCTCCCTTAAATAAATAATAA 1821
DB 68642 TGCCATTGCACTCCAGCTTGGGCGACAGAGCGAGCTCTCTCTCAAAAAAATAAAAAA 68583
QY 1822 TTTTAAAAAATAGGGGTACTAATATCTACCTTAAAGGATGAGGGTTAAATTAAGTAC 1881
DB 68582 AAAAAAAGCAGAGTCTGGAACAGTGCACATAGTATGCCAGTTTGTGTTTAAAGA 68523
QY 1882 ACACATAAGCCCTAGCGCAGTGGCTTATGCTGTGTAATCTCAACACTTTGGAGTCTGTGG 1941
DB 68522 AAAGTGTAGGCTGGATGCAATGGCTCGCACCCTACAGTCCAGCACTTTGGGAGCGGAGG 68463
QY 1942 CGGAGGATCACTTGACCCAGGAGTTTGAGACTAGTCTGGGCAACA----GAGACATGT 1997
DB 68462 CAGAAGGATTTGCTTTAGGCGAGGAGTTCAAGACAGCGCTGGCAACATAGCGGAACCTG 68403
QY 1998 CTCTATATGTTGTTTGTGTTTGTGTTTACAGGTGTTGTTGTCAGCTGCAGTCCCA 2057
DB 68402 TTCTATGAAGATAAATTAATAAATTAGCAGCGCTGTTGTTGTCAGCTGTAGTCCCA 68343
QY 2058 GCTACTAGGAGGCTGAGGTGGGAGGACTGCTGAGCGCCAGGAGGTGAGGCTGCAGTGA 2117
DB 68342 GCTACTTGGGAGGCTGTTGGTAGGAGGATGGCTTGAGTCTAGGAGTTTGGGCTGCAGTGA 68283
QY 2118 GCGATGATTGCGCACTGCACTCCAGCTCGGCGCAACACAGCAGACCTTGTCTCAAAAAC 2177
DB 68282 GCTATGATGGCGCGCTGCACTCCAGCTTGGGCAACACAGTGAACCTGTCTCTAAAT 68223
QY 2178 AAACAAAAAGCATACTCATATA 2197

DB 68222 AAAATAAAGGTTTTTTTAA 68203
RESULT 15
ADRI6283/c
ID ADRI6283 standard; DNA; 47219 BP.
XX
XX AC ADRI6283;
XX
XX DT 21-OCT-2004 (first entry)
XX
XX DE Human Cayman ataxia splice site mutant DNA.
XX
XX KW Human; jittery; de; Cayman ataxia; ATCAV; KIAA1872; chromosome 19p13.3;
KW ataxia; myoclonus; dystonia; epilepsy; nyctagmus; splice site mutant;
KW SNP; single nucleotide polymorphism.
XX
XX OS Homo sapiens.
XX
XX FH Key Location/Qualifiers
FT misc_feature 33181..33190
FT /tag= a
FT /note= "Exon 9-intron 9 boundary"
FT variation replace(33187,G)
FT /tag= b
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "Causes splice site mutation"
XX
XX US2004146900-A1.
XX
XX PN 29-JUL-2004.
XX
XX PD 03-NOV-2003; 2003US-00699941.
XX
XX PF 01-NOV-2002; 2002US-0422971P.
XX
XX PR 08-NOV-2002; 2002US-0424973P.
XX
XX PA (UNMI) UNIV MICHIGAN.
XX
XX PI Burmeister M;
XX
XX WPI; 2004-552665/53.
XX
XX Detecting variant Cayman ataxia polypeptide or nucleic acid sequence in
XX subject, by detecting presence or absence of variant Cayman ataxia
XX polypeptide or nucleic acid in biological sample.
XX
XX Claim 5; SEQ ID NO 10; 98pp; English.
XX
XX The invention relates to detecting variant Cayman ataxia polypeptide or
XX nucleic acid sequence (ATCAV) in a subject, involving providing a
XX biological sample from a subject, where the biological sample comprises a
XX Cayman ataxia polypeptide or nucleic acid, and detecting the presence or
XX absence of a variant Cayman ataxia polypeptide or nucleic acid in the
XX biological sample. Also included is a kit comprising a reagent for
XX detecting the presence or absence of a variant Cayman ataxia nucleic acid
XX or polypeptide in a biological sample. The human ATCAV gene (previously
XX isolated as cDNA KIAA1872, located on chromosome 19p13.3) was identified
XX and mapped using the fact that it is the homologue of the mouse jittery
XX gene, located on mouse chromosome 10. The method is useful for detecting
XX variant Cayman ataxia polypeptide or nucleic acid sequence in a subject.
XX The presence of the variant Cayman ataxia polypeptide or nucleic acid is
XX indicative of a disorder chosen from ataxia, myoclonus, dystonia,
XX epilepsy, and nyctagmus in the subject. The biological sample is chosen
XX from blood sample, a tissue sample, urine sample, saliva sample, and an
XX amniotic fluid sample. The subject is chosen from embryo, foetus, newborn
XX animal, young animal, and an adult animal. The animal is a human. The
XX human is an adult female of child-bearing age. The present sequence is
XX the DNA representing a Human ATCAV splice site mutant (disease causing)
XX where a single nucleotide polymorphism in the splice site of exon 9-
XX intron 9 causes a read through mutant protein to be produced.

```

SQ Sequence 47219 BP; 11562 A; 12203 C; 12045 G; 11209 T; 0 U; 200 Other;

Query Match      13.0%; Score 321.4; DB 13; Length 47219;
Best Local Similarity 68.9%; Pred. NO. 7.1e-61;
Matches 513; Conservative 0; Mismatches 226; Indels 6; Gaps 5;

QY 1461 TTTTGGACGAGTTGTTTAACTCTTTGTACCTCAGCTCTTCTTATACAAAAATAGGGTA 1520
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DY 20591 TCTTGTGCGAGAGAATGGNAATAGTTTATGTACAGTTCTGCCCTCAAGNAGATGAAGA 20532

QY 1521 CTAGCCAGCGGGTGGCTGGCCCTGTATATCCAGACACTTGGGAGGCGGAGCGCCG 1580
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DY 20531 TAGGCCGGGTGGCTGGCTGATATCCAGCACTTTGGAGGCGGAGCGCTGCG 20472

QY 1581 GATCACTTGAGGTCAGAGTTTCAGACCACTGGTCAACATGGGTGAAACCGCGCTCT 1640
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DY 20471 GATCAC - GAGGTCAGAGATCGAGACCATCTCTGGCTAACAT - GGTGAAACCCCGTCTCT 20415

QY 1641 ACCAAAAATA - TAAAACTTAGCTGAGTGTGTAGCGCATGACTGTAAATCCCAAGCAACTC 1699
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DY 20414 ACTTAAAGTACAAAAAATAAGCGGGCGTGTGGCAGGTGCTGTAGTCCAGCTACTC 20355

QY 1700 AGGAGCTGAGGCA - GAGATCGCTTTGAACCTGGGAGGCGAGGTTGCAGTGAGCTGAGA 1758
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DY 20354 AGGAGCTGAGGCGAGGAGATGCAATGAACCCAGGAGGCGAGCTTGCAGTGAGCGAGA 20295

QY 1759 TCGTACCACTGCACTCAGCTTGGGCGACAGAGCGAGACTCTGCCCTTAAAAATAATAA 1818
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DY 20294 TCGTGCACCTGCACTCAGCTTGGGCGACAGAGCAAGACTCTGTCTCAAAAAAATAAAA 20235

QY 1819 TAATTTTAAAAAATAAGGGTACTAAATCTACCTTAAAGGATGAGGTTAAATTAAG 1878
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DY 20234 AAAAAAATAAGATGAGGATACTCTGTGTAGGCTGTGTATAGTATGTCTATCCAAA 20175

QY 1879 TACACACATAAGCCCTAGCGAGTGGCTTATGCTGTATCTCAACACTTTGGGAGTCTG 1938
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DY 20174 AATACAGGATGGCCAGGCGCAGAGCTCACGCTGTGTATCCAGCACTTTGGGAGGCTG 20115

QY 1939 TGGGGGAGGATCACTTGAGCCCGAGAGTTTGAGACTAGTCTGGGCAACAGACATGTC 1998
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DY 20114 AGGAGGAGGATCACTTGAGCTCAGGAATTCGAGACCAAGCTGGCAACATGTTGAAACG 20055

QY 1999 TCTATAGTTGTGTGTTTGTGTTTACAGGTGTGGTGTGACCTGCACTGCCAG 2058
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DY 20054 CCATCTCTAAAAATAACAAAAATTAGCGAGGATGGTGGAGGACACCTGTATATCCAG 19995

QY 2059 CTACTAGGAGGCTGAGGTGGAGGACTGCTGAGCCCGAGGAGGTGAGGCTCAGTGAG 2118
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DY 19994 CTA - TCAGGAGGCTGAGGCGAGGAGATCGCTTGAACCCGGGAGGCGGAGTTGAGTGAG 19936

QY 2119 CCATGATTGTGCCACTGCACTCAGCTTGGGCAACACAGCAAGCACTTGTCTCAAAAACA 2178
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DY 19935 CCAGATCTGTGCCACTGCTCTCCAGCTGGGTGACAGAGTGAGACTCCATCTCAAAAACA 19876

QY 2179 AACAAAAAGCATACTCATAGTGC 2203
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DY 19875 ACAACAAACAAACTATCAAGGC 19851

```

Search completed: August 11, 2005, 23:45:38
Job time : 2326 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 11, 2005, 22:35:08 ; Search time 10557 Seconds

(without alignments)
11332.379 Million cell updates/sec

Title: US-10-829-118-2

Perfect score: 2469

Sequence: 1 ttgagtcgtctcgaggctcc.....gaatacatggccctacagct 2469

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2469	100.0	2469	6	AR527528 Sequence
2	2447.4	99.1	212596	9	AC040169 Homo sapi
3	2447.4	99.1	216522	2	AC139664 Homo sapi
4	980	39.7	980	6	AR527527 Sequence
5	521.4	21.1	532	9	BC052334 Homo sapi
6	347	14.1	148864	9	Z77249 Human DNA s
7	346	14.0	46741	9	AC010519 Homo sapi
8	344	13.9	194241	2	AC093324 Homo sapi
9	343.4	13.9	167587	9	AC108000 Homo sapi
10	339.4	13.7	52479	2	AC074125 Homo sapi
11	339.4	13.7	86684	9	AC010605 Homo sapi
12	339.4	13.7	169583	9	AC080005 Homo sapi
13	339	13.7	17992	2	AC020715 Homo sapi
14	338.4	13.7	173252	9	AC036176 Homo sapi
15	338.4	13.7	188923	9	AC009267 Homo sapi
16	338.2	13.7	231160	2	AC135717 Homo sapi
17	337	13.6	185624	9	AC103923 Homo sapi
18	336.4	13.6	43457	9	AL499628 Human DNA
19	336.4	13.6	81579	9	AE006640 Homo sapi

20	336.4	13.6	110000	2	BX324168_4	Continuation (5 of
21	335.8	13.6	109488	9	HS298J15	AL031774 Human DNA
22	335	13.6	158800	2	AC027090	AC027090 Homo sapi
23	335	13.6	192096	2	AL590133	AL590133 Human DNA
24	334.8	13.6	221357	2	AP004246	AP004246 Homo sapi
25	334.4	13.5	201659	2	AC146419	AC146419 Pan trogl
26	334.2	13.5	184393	9	AC018665	AC018665 Homo sapi
27	333.8	13.5	4124	6	AX098245	AX098245 Sequence
28	333.8	13.5	110293	9	HS121G12	AL049766 Human DNA
29	333.4	13.5	105230	9	HS121G12	AL109965 Human DNA
30	332.8	13.5	160979	2	AC055879	AC055879 Homo sapi
31	332.2	13.5	211063	2	AC069254	AC069254 Homo sapi
32	332	13.4	97120	9	AP004147	AP004147 Homo sapi
33	332	13.4	34980	6	CQ869981	CQ869981 Sequence
34	331.4	13.4	50417	9	AY497474	AY497474 Homo sapi
35	331.4	13.4	101624	9	AC104116	AC104116 Homo sapi
36	331.4	13.4	160570	2	AC092317	AC092317 Homo sapi
37	330.8	13.4	157871	9	AC104365	AC104365 Homo sapi
38	330.6	13.4	123501	9	AL390241	AL390241 Human DNA
39	330.6	13.4	145166	2	AC053473	AC053473 Homo sapi
40	330.6	13.4	164810	2	AC023411	AC023411 Homo sapi
41	330.6	13.4	266079	2	AL391060	AL391060 Homo sapi
42	330.2	13.4	163338	9	AL162426	AL162426 Human DNA
43	330.2	13.4	172588	2	AC007430	AC007430 Homo sapi
44	330	13.4	33026	9	AF156673	AF156673 Homo sapi
45	330	13.4	177893	9	AC005089	AC005089 Homo sapi

ALIGNMENTS

RESULT 1	AR527528	Sequence 2 from patent US 6723553.	2469 bp	DNA	linear	PAT 08-OCT-2004
LOCUS	AR527528	Sequence 2 from patent US 6723553.	2469 bp	DNA	linear	PAT 08-OCT-2004
DEFINITION	AR527528	Sequence 2 from patent US 6723553.	2469 bp	DNA	linear	PAT 08-OCT-2004
ACCESSION	AR527528	Sequence 2 from patent US 6723553.	2469 bp	DNA	linear	PAT 08-OCT-2004
VERSION	AR527528.1	GI:53914633	2469 bp	DNA	linear	PAT 08-OCT-2004
KEYWORDS	Unknown.					
SOURCE	Unknown.					
ORGANISM	Unknown.					
REFERENCE	1 (bases 1 to 2469)					
AUTHORS	Abrahamsen, L., Ekblom, J., Forsgren, M., Horling, J. and Johansson, P.					
TITLE	Promoter sequences					
JOURNAL	Patent: US 6723553-A 2 20-APR-2004;					
FEATURES	Location/Qualifiers					
source	1..2469					
	/organism="unknown"					
	/mol_type="genomic DNA"					
ORIGIN						

Query Match	100.0%	Score 2469;	DB 6;	Length 2469;
Best Local Similarity	100.0%;	Pred. No. 0;	Mismatches 0;	Indels 0;
Matches 2469;	Conservative 0;			
Qy	1	TTGAGTCTCTCGAGGCTCCGGCCAGACGAGCGGTATTGTTTTCACTCGGTGAATGCT	60	
Db	1	TTGAGTCTCTCGAGGCTCCGGCCAGACGAGCGGTATTGTTTTCACTCGGTGAATGCT	60	
Qy	61	CATTTTCACCTAAAGAAAAACAGGCAACGAACTGCGGAGCGGAGACCCCGCA	120	
Db	61	CATTTTCACCTAAAGAAAAACAGGCAACGAACTGCGGAGCGGAGACCCCGCA	120	
Qy	121	GGGCGCGGTACAGGACCGCTGTGTCCAAACAGCGCGGAGCGCCCGCCACCTCCC	180	
Db	121	GGGCGCGGTACAGGACCGCTGTGTCCAAACAGCGCGGAGCGCCCGCCACCTCCC	180	
Qy	181	CGGACCGCGCGCGCGCGCGCGCGCTCGCTCGGCGCTCGGACGCAACCGCACCT	240	
Db	181	CGGACCGCGCGCGCGCGCGCGCGCTCGCTCGGCGCTCGGACGCAACCGCACCT	240	
Qy	241	GAGCGAGCGCGCGCGCGCGCGCGCTAGGCGGAGCGCGCGCGCGCGCGCTG	300	
Db	241	GAGCGAGCGCGCGCGCGCGCGCGCTAGGCGGAGCGCGCGCGCGCGCGCTG	300	

Db 241 GAGCAGCGGCGCCACCGCTAGGCGAGCGGCTCGGGGAGGCGCGCGCGCGGCTG 300
Qy 301 ACGTACTTGGCCCGCGGGAGCTCAGGGCCGGCGGGCCCGGGATGACGGGGCCCTCCGGG 360
Db 301 ACGTACTTGGCCCGCGGGAGCTCAGGGCCGGCGGGCCCGGGATGACGGGGCCCTCCGGG 360
Qy 361 CGAACACGCTGGGCACTCCATTCGGGGCTGTGTTTACTCCCAACTCTCGCGAGACTGGGG 420
Db 361 CGAACACGCTGGGCACTCCATTCGGGGCTGTGTTTACTCCCAACTCTCGCGAGACTGGGG 420
Qy 421 GCCGGGCCAGGAGGCCACAGCTGGGAGCCTCAGCTCCGCCACCCAGCGTCCCTGTG 480
Db 421 GCCGGGCCAGGAGGCCACAGCTGGGAGCCTCAGCTCCGCCACCCAGCGTCCCTGTG 480
Qy 481 TGTCCCGCGCTCCCGGGCTTGGTGGCGGCTCTGGAACGCGCTGCGGAGCGGGACACGCG 540
Db 481 TGTCCCGCGCTCCCGGGCTTGGTGGCGGCTCTGGAACGCGCTGCGGAGCGGGACACGCG 540
Qy 541 CGGGAGGATGGACGAAGGTGCTCGCACATTTGCGCGCGGGGGCGGGTGGCAGGGTG 600
Db 541 CGGGAGGATGGACGAAGGTGCTCGCACATTTGCGCGCGGGGGCGGGTGGCAGGGTG 600
Qy 601 AAGCGAGGGCGTGGCCAGCAGCTGCCAGCGCGGAGAACGCGCTGGGGGAACCTTG 660
Db 601 AAGCGAGGGCGTGGCCAGCAGCTGCCAGCGCGGAGAACGCGCTGGGGGAACCTTG 660
Qy 661 GTCCGCTCTGCGGCTCGCTTAGGATCCCGGAAGAGGACCGGGCGCGGAAGCGGCCAG 720
Db 661 GTCCGCTCTGCGGCTCGCTTAGGATCCCGGAAGAGGACCGGGCGCGGAAGCGGCCAG 720
Qy 721 GCTGGGCGAGGATCAGAAAGCTGCTGCGCAGGCTCCCTGCCCGCGGCGCTGCTG 780
Db 721 GCTGGGCGAGGATCAGAAAGCTGCTGCGCAGGCTCCCTGCCCGCGGCGCTGCTG 780
Qy 781 TCATGGAAGCTCGTGAGAGCTCGCTTCCCGCGGAGCCCTTCTGCGAGGGTCCACGTCCA 840
Db 781 TCATGGAAGCTCGTGAGAGCTCGCTTCCCGCGGAGCCCTTCTGCGAGGGTCCACGTCCA 840
Qy 841 GGCACCGGGCTCGGACACCCACCGCGCGGGGACCTGCGCTGGGTGCCCTTTAAC 900
Db 841 GGCACCGGGCTCGGACACCCACCGCGCGGGGACCTGCGCTGGGTGCCCTTTAAC 900
Qy 901 CCGGGCGGTAGCTCGTTAGATGGCAAGTCTCCGGTCCGGACACCGCAACCCCAAT 960
Db 901 CCGGGCGGTAGCTCGTTAGATGGCAAGTCTCCGGTCCGGACACCGCAACCCCAAT 960
Qy 961 CCGGCTGCCGACCTCTGACCCCGGCCACCGGACGACAGACTGGGCTCCCGAGC 1020
Db 961 CCGGCTGCCGACCTCTGACCCCGGCCACCGGACGACAGACTGGGCTCCCGAGC 1020
Qy 1021 CGAGCGGCTGCCGACACCGGTCGTCGGAACCGGAGGACCTTGTAAACGCGACGTCG 1080
Db 1021 CGAGCGGCTGCCGACACCGGTCGTCGGAACCGGAGGACCTTGTAAACGCGACGTCG 1080
Qy 1081 TTTGCTCTTTTGAAGAAACAGATAAATGTCTAACTGTCTGAAAAGCTTGGCCCT 1140
Db 1081 TTTGCTCTTTTGAAGAAACAGATAAATGTCTAACTGTCTGAAAAGCTTGGCCCT 1140
Qy 1141 AAAAGATGTCTGGGTGACTCTAGATGCTAGGATCAGTTTGTGTTTCAATGTAATGGACCA 1200
Db 1141 AAAAGATGTCTGGGTGACTCTAGATGCTAGGATCAGTTTGTGTTTCAATGTAATGGACCA 1200
Qy 1201 CCGGAGCTCGGTACGGACACCGGTGTCGGAACCGGAGGACCTTGTAAACGCGACGTCG 1260
Db 1201 CCGGAGCTCGGTACGGACACCGGTGTCGGAACCGGAGGACCTTGTAAACGCGACGTCG 1260
Qy 1261 GGTGATGCTACAGCGCGATCAGACAGTTTGTCTTCTGAACTTGACACTGCACCA 1320
Db 1261 GGTGATGCTACAGCGCGATCAGACAGTTTGTCTTCTGAACTTGACACTGCACCA 1320
Qy 1321 CGGTAATGCTGAATGACCAATATTTACAGATCAGCGCATCTTCTTCAACATGA 1380
Db 1321 CGGTAATGCTGAATGACCAATATTTACAGATCAGCGCATCTTCTTCAACATGA 1380

Qy 1381 TTTAACACAGTTGACTTAATATGTTGAATAGTAGATCACAATTTACCATACCCAC 1440
Db 1381 TTTAACACAGTTGACTTAATATGTTGAATAGTAGATCACAATTTACCATACCCAC 1440
Qy 1441 CTCAGGCTTCTACTTTCGTAATTTTTCAGCAGGTTGTTTAACTCTTTGTACCTCAGCTTCT 1500
Db 1441 CTCAGGCTTCTACTTTCGTAATTTTTCAGCAGGTTGTTTAACTCTTTGTACCTCAGCTTCT 1500
Qy 1501 TCATTACAAAAATAGGGTACTAGCAGCGCGGGTGGCTCGCGCTGTGTAATCCAGCACT 1560
Db 1501 TCATTACAAAAATAGGGTACTAGCAGCGCGGGTGGCTCGCGCTGTGTAATCCAGCACT 1560
Qy 1561 TGGGAGGCGGAGCAGCCGATCATTGAGCTCAGAGTTTCAGACCGAGCTGGTCAAC 1620
Db 1561 TGGGAGGCGGAGCAGCCGATCATTGAGCTCAGAGTTTCAGACCGAGCTGGTCAAC 1620
Qy 1621 ATGGGTGAAACCGCCGCTCTACAAAAATATAAAAACTTTAGCTGAGTGTGGTAGCGCATG 1680
Db 1621 ATGGGTGAAACCGCCGCTCTACAAAAATATAAAAACTTTAGCTGAGTGTGGTAGCGCATG 1680
Qy 1681 ACTGTAATCCCAAGCAACTCAGGAGCTGAGGAGAGAAATCGCTTGAACTCTGGAGGCGGA 1740
Db 1681 ACTGTAATCCCAAGCAACTCAGGAGCTGAGGAGAGAAATCGCTTGAACTCTGGAGGCGGA 1740
Qy 1741 GGTTCGAGTGTGAGTGTGATCGTACCACTCCAGCTTGGGCGACAGAGCGAGCTCT 1800
Db 1741 GGTTCGAGTGTGAGTGTGATCGTACCACTCCAGCTTGGGCGACAGAGCGAGCTCT 1800
Qy 1801 GCCTTAAAAATATAATAAATTTTAAAAAAATAGGGGTACTAATATCTACCTTTAAAG 1860
Db 1801 GCCTTAAAAATATAATAAATTTTAAAAAAATAGGGGTACTAATATCTACCTTTAAAG 1860
Qy 1861 GATGAGGTTAAATTAAGTACACATAAGCCCTAGCGAGTGGCTTATGCTGTAATCT 1920
Db 1861 GATGAGGTTAAATTAAGTACACATAAGCCCTAGCGAGTGGCTTATGCTGTAATCT 1920
Qy 1921 CAACACTTTGGGAGTCTGTGGCGGAGGATCACTTTAGCCCGAGGAGTTTGAGACTAGTCT 1980
Db 1921 CAACACTTTGGGAGTCTGTGGCGGAGGATCACTTTAGCCCGAGGAGTTTGAGACTAGTCT 1980
Qy 1981 GGGCAACAGACATGTCTCTATAGTTGTGTTGTTTGTGTTTGTGTTTGTGTTTGTGTT 2040
Db 1981 GGGCAACAGACATGTCTCTATAGTTGTGTTGTTTGTGTTTGTGTTTGTGTTTGTGTT 2040
Qy 2041 GTGCACTGCACTCCAGCTACTAGGAGGCTGAGGTGGGAGGACTGCTGAGCCCGAG 2100
Db 2041 GTGCACTGCACTCCAGCTACTAGGAGGCTGAGGTGGGAGGACTGCTGAGCCCGAG 2100
Qy 2101 GGTGAGGCTGAGTGGCCATGATTTGCGCACTGCACCTCCAGCTGGGCAACACAGCAA 2160
Db 2101 GGTGAGGCTGAGTGGCCATGATTTGCGCACTGCACCTCCAGCTGGGCAACACAGCAA 2160
Qy 2161 GACCTTGTCTCAAAACAAACAAAGCATACTCAATAAGTGTCTGGCTCTCTATATGAT 2220
Db 2161 GACCTTGTCTCAAAACAAACAAAGCATACTCAATAAGTGTCTGGCTCTCTATATGAT 2220
Qy 2221 CAATATGCTGGTGGATTTCTTGAATCTTCTGACTGAGTCTCATACGATTTTCTGA 2280
Db 2221 CAATATGCTGGTGGATTTCTTGAATCTTCTGACTGAGTCTCATACGATTTTCTGA 2280
Qy 2281 ACTTTTGGAGAACTCTTGGCTCTCTGCAATTTGCAACCGCTCAAGGCACTCCCTCTGCC 2340
Db 2281 ACTTTTGGAGAACTCTTGGCTCTCTGCAATTTGCAACCGCTCAAGGCACTCCCTCTGCC 2340
Qy 2341 ACCACACAAGCATTTGATTTTAACTTGTCTTCTTCTGTTCCAACTTTAGGTAAA 2400
Db 2341 ACCACACAAGCATTTGATTTTAACTTGTCTTCTTCTGTTCCAACTTTAGGTAAA 2400
Qy 2401 TTAATCTTGGTCAGGGTCTCTGAAACAGCCCTTTTAGTCACTATGCCATTTGAATACATGGC 2460
Db 2401 TTAATCTTGGTCAGGGTCTCTGAAACAGCCCTTTTAGTCACTATGCCATTTGAATACATGGC 2460

Db	174214	GGTGAATGCGCTACAGGCGGATCAGACAGCTTTTGTGTCTTTTGGAACTTGACACTGCAACA	174273
Qy	1321	CGGTAAATGCTGAATGCAACCAATATTTACAGATCAGAGCGCATCATCTTCTTCAACATGA	1380
Db	174274	CGGTAAATGCTGAATGCAACCAATATTTACAGATCAGAGCGCATCATCTTCTTCAACATGA	174333
Qy	1381	TTTAAACACAGTTGACTTAATATGTTGGATAAATGTAGAAATCAAAATTAACCATACCCAC	1440
Db	174334	TTTAAACACAGTTGACTTAATATGTTGGATAAATGTAGAAATCAAAATTAACCATACCCAC	174393
Qy	1441	CTCAGGCTTCTACTCTTGGTAAATTTTGAACAGGTGTGTTTAACTCTTTTGTACCTGATCT	1500
Db	174394	CTCAGGCTTCTACTCTTGGTAAATTTTGAACAGGTGTGTTTAACTCTTTTGTACCTGATCT	174453
Qy	1501	TCATTTACAAAATAGGGGTACTAGCCAGCGGGGTGGCTCGCGCTGTAAATCCCAGCAT	1560
Db	174454	TCATTTACAAAATAGGGGTACTAGCCAGCGGGGTGGCTCGCGCTGTAAATCCCAGCAT	174513
Qy	1561	TGGGAGGCGGAGGCGCGGATCACTTTGAGGTCAAGAGTTTTCAGACACGAGCTGGTCAAC	1620
Db	174514	TGGGAGGCGGAGGCGCGGATCACTTTGAGGTCAAGAGTTTTCAGACACGAGCTGGTCAAC	174573
Qy	1621	ATGGGTGAAACCGCGGCTCTACCAAAAATATAAAACTTTAGCTGAGTGTGGTAGCGCATG	1680
Db	174574	ATGGGTGAAACCGCGGCTCTACCAAAAATATAAAACTTTAGCTGAGTGTGGTAGCGCATG	174633
Qy	1681	ACTGTAAATCCGAGCAATCTCAGGAGGCTGAGGAGAGAAATCGCTTGAACCTGGGAGCGGA	1740
Db	174634	ACTGTAAATCCGAGCAATCTCAGGAGGCTGAGGAGAGAAATCGCTTGAACCTGGGAGCGGA	174693
Qy	1741	GGTTGCAGTGAAGTCTGAGATCGTACCACCTGCACTCTCAGAGCTTGGCGCAGACGAGAGTCT	1800
Db	174694	GGTTGCAGTGAAGTCTGAGATCGTACCACCTGCACTCTCAGAGCTTGGCGCAGACGAGAGTCT	174753
Qy	1801	GCCTTAAAAATAAAATAAATAAATTTTTTAAAAAAAATAGGGGTACTTAATATCTACCTTAAAG	1860
Db	174754	GCCTTAAAAATAAATAAATAAATTTTTTAAAAAAAATAGGGGTACTTAATATCTACCTTAAAG	174813
Qy	1861	GATGAGGGTTAAATTAAGTACACATAAGCCCTAGCGAGTGGCTTATGCCCTGTAATCT	1920
Db	174814	GATGAGGGTTAAATTAAGTACACATAAGCCCTAGCGAGTGGCTTATGCCCTGTAATCT	174873
Qy	1921	CAACACTTTGGAGTCTGTGGCGGAGGATCACTTTGAGCCAGGAGTTTGAAGTACTAGTCT	1980
Db	174874	CAACACTTTGGAGTCTGTGGCGGAGGATCACTTTGAGCCAGGAGTTTGAAGTACTAGTCT	174933
Qy	1981	GGGCAACAGACATGTCTCTATAGTTGTGTTTGGTTTGTGTTTACCAAGTGTGGTGGT	2040
Db	174934	GGGCAACAGACATGTCTCTATAGTTGTGTTTGGTTTGTGTTTACCAAGTGTGGTGGT	174993
Qy	2041	GTGCACCTGCAGTCCAGCTACTAGGGAGGCTGAGGTGGGAGGATGTCGCTGAGCCACAGGA	2100
Db	174994	GTGCACCTGCAGTCCAGCTACTAGGGAGGCTGAGGTGGGAGGATGTCGCTGAGCCACAGGA	175053
Qy	2101	GGTCGAGGCTGCAGTGAGCCATGATTTGGCCACTGCACCTCCAGCCTGGGCAACACAGCAA	2160
Db	175054	GGTCGAGGCTGCAGTGAGCCATGATTTGGCCACTGCACCTCCAGCCTGGGCAACACAGCAA	175113
Qy	2161	GACCTTGTCTCAAAAACAAAACAAAAGCATCTCATAAAAGTGTCTGGCTCTCTATATGAT	2220
Db	175114	GACCTTGTCTCAAAAACAAAACAAAAGCATCTCATAAAAGTGTCTGGCTCTCTATATGAT	175173
Qy	2221	CAATATGTTGGTGGATTTCTTTGAATCCCTTCTGACTCAGATCTCATACGATTTTCTGA	2280
Db	175174	CAATATGTTGGTGGATTTCTTTGAATCCCTTCTGACTCAGATCTCATACGATTTTCTGA	175233
Qy	2281	ACTTTTGGAGAACTCTTGGCTCTCTGCAATTTGCAAAACCGTCAAAAGCATCTCCCTCTGCC	2340
Db	175234	ACTTTTGGAGAACTCTTGGCTCTCTGCAATTTGCAAAACCGTCAAAAGCATCTCCCTCTGCC	175293
Qy	2341	ACCACACAAAGCATTTGATTTTTTAAACTTGACTATGTCTTCTTGTCTTGTCCAACTTTAGGTAAA	2400

Db	175294	ACCACACAAAGCAATTGATTTAAACTTTGACATATGTCTCTGTTCCAACTTTAGGTAA	175353
Qy	2401	TTAATCTTTGGTCAGGGTTCTCTGAACAGCCCTTTAGTCACTATGCAATTTGAATACATGCG	2460
Db	175354	TTAATCTTTGGTCAGGGTTCTCTGAACAGCCCTTTAGTCACTATGCAATTTGAATACATGCG	175413
Qy	2461	CCTACAGCT 2469 	
Db	175414	CCTACAGCT 175422	
RESULT 3			
AC139664/c			
LOCUS			
DEFINITION	AC139664	216522 bp DNA linear HTG 27-MAR-2003	
		Hom sapiens clone RP11-812D13, WORKING DRAFT SEQUENCE, 9 unordered	
		pieces.	
ACCESSION	AC139664		
VERSION	AC139664.5	GI:29293970	
KEYWORDS	HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_FULLTOP.		
SOURCE	Hom sapiens (human)		
ORGANISM	Hom sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.		
	M1 (bases 1 to 216522)		
	Muzny D.M., Adams C., Adio-Oduola B., Ali-osman F.R., Allen C.,		
	Alabrooks S.L., Amaratunge H.C., Are J.R., Ayelle M., Banks T.,		
	Bourcia J., Benton J., Bivaga M., Brown E., Brown M., Bryant N.P.,		
	Buhay C., Burch P., Burkett C., Burrell K.L., Byrd N.C.,		
	Carroll T.F., Carter M., Cavazos S.R., Chacko J., Chavez D.,		
	Chen G., Chen R., Chen Z., Chowdhry I., Christopoulos C.,		
	Cleveland C.D., Cox C., Coyle M.D., Dathorne S.R., David R.,		
	Davila M.L., Davis C., Davy-Carroll L., Dederich D.A.,		
	Delaney K.R., Delgado O., Dunn A.L., Ding Y., Dinh H.H.,		
	Dolanwaite K.J., Draper H., Dugan-Rocha S., Durbin K.J.,		
	Earnhart C., Edgar D., Edwards C.C., Elhaj C., Escotto M.,		
	Falls T., Ferraguto D., Flagg N., Ford J., Foster P., Frantz P.,		
	Gabisi A., Gao J., Garcia A., Garner T., Garza N., Gill R.,		
	Gratelli J.H., Guevara W., Gunaratne P., Hale S., Hamilton K.,		
	Harris C., Harris K., Hart M., Havlak P., Hawes A., Hernandez J.,		
	Hernandez O., Hodgson A., Hogue M., Holloway C., Hollins B.,		
	Homsif F., Howard S., Huber J., Hulyk S., Hume J., Jackson L.E.,		
	Jacobson B., Jia Y., Johnson R., Jolivet S., Joudah S.,		
	Karlsson E., Kelly S., Khan U., King L., Korvah J., Kovar C.,		
	Kratovic J., Kureshi A., Landry N., Leal B., Lewis L.C., Lewis L.,		
	Li J., Li Z., Licharge O., Lieu C., Liu J., Liu W., Louiseged H.,		
	Lozano R.J., Lu X., Lucier A., Lucier R., Luna R., Ma J.,		
	Maheshwari M., Mapua P., Martin R., Martindale A., Martinez E.,		
	Masshey E., Mawhinney E., McLeod M.P., Meador M., Mei G., Metzker M.,		
	Miner G., Miner Z., Mitchell T., Mohabbat K., Morgan M., Morris S.,		
	Moser M., Neal D., Newton J., Newton N., Nguyen A., Nguyen N.,		
	Nguyen N., Nickerson E., Nwokenwo S., Ogum H., Okwuonu G.,		
	Oragunye N., Oviedo R., Pace A., Payton B., Peery J., Perez L.,		
	Peters L., Pickens R., Primus E., Pu L.L., Quiles M., Ren Y.,		
	Rives M., Rojas A., Rojibokan I., Rolfe M., Ruiz S., Savary G.,		
	Scherz S., Scott G., Shen H., Shoohtari N., Sisson I.,		
	Sodergren E., Sonake T., Sparks A., Stanley H., Stone H.,		
	Sutton A., Svatek A., Tabor P., Tamerisa A., Tamerisa K., Tang H.,		
	Tansey J., Taylor C., Taylor T., Telford B., Thomas N., Thomas S.,		
	Umanik K., Vasquez L., Vera V., Villalon D., Vinson R., Wang Q.,		
	Wang S., Ward-Moore S., Warren R., Washington C., Watlington S.,		
	Williams G., Williamson A., Wleczek R., Wooden S., Worley K.,		
	Wu C., Wu Y.F., Zhou J., Zorrilla S., Nelson D.,		
	Weinstock G. and Gibbs R.		
	Direct Submission		
	Unpublished		
	2 (bases 1 to 216522)		
	Worley K.C.		
	Direct Submission		
	Submitted (08-FEB-2003) Human Genome Sequencing Center, Department		
	of Molecular and Human Genetics, Baylor College of Medicine, One		
	Baylor Plaza, Houston, TX 77030, USA		
	3 (bases 1 to 216522)		
REFERENCE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			

AUTHORS TITLE JOURNAL

Worley, K. C.
Direct Submission
Submitted (27-MAR-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Mar 27, 2003 this sequence version replaced gi:28467047.

COMMENT

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Drafting Center Code: BCM
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HET1
Center clone name: RP11-812D13
----- Summary Statistics

Sequencing vector: Plasmid;
Chemistry: Dye-terminator; Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 215023 bases at least Q40
Consensus quality: 215949 bases at least Q30
Consensus quality: 216649 bases at least Q20
Estimated insert size: 210467; sum-of-contrigs estimation
Quality coverage: 10x in Q20 bases; sum-of-contrigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contrigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contrigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1056: contig of 1056 bp in length
* 1057 1156: gap of unknown length
* 1157 2298: contig of 1142 bp in length
* 2299 2398: gap of unknown length
* 2399 3437: contig of 1039 bp in length
* 3438 3537: gap of unknown length
* 3538 5421: contig of 1884 bp in length
* 5422 5521: gap of unknown length
* 5522 6548: contig of 1027 bp in length
* 6549 7859: contig of 1211 bp in length
* 7860 7959: gap of unknown length
* 7960 21513: contig of 13554 bp in length
* 21514 21613: gap of unknown length
* 21614 40787: contig of 19174 bp in length
* 40788 40888: gap of unknown length
* 40889 216522: contig of 175635 bp in length.

FEATURES source

1. 216522
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="RP11-812D13"

ORIGIN

Query Match 99.1%; Score 2447.4; DB 2; Length 216522;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 2462; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

Qy 1 TTGAGTCTGTCTGGAGCTCCGGCCAGACGAGCGGCGTATTGTTTCACTCGGTGAATGCT 60
Db 19780 TTGAGTCTGTCTGGAGCTCCGGCCAGACGAGCGGCGTATTGTTTCACTCGGTGAATGCT 19721

Qy 61 CATTTACGTAAGAAACACGAGCAACGAACTGCCGAGCGCGGACGCCCGCA 120
Db 19720 CATTTACGTAAGAAACACGAGCAACGAACTGCCGAGCGCGGACGCCCGCA 19661

Qy 121 GGGCCGGGTACAGCAGCTGTGTCCAAACAGCGCGGAGGGCCCGCCACCTCCC 180

19660	GGGCGCGGTACAGCAGCTGTGTGTCCAAACAGCGCGGAGGCCCGCGCCACTCCC	19601
181	CCGACCCGGCCCGCCCGCCCGAGCCCTCGCTCGGGGGCTCGAGCGCAACCGGCACACT	240
19600	CCGACCCGGCCCGCCCGCCCGAGCCCTCGCTCGGGGGCTCGAGCGCAACCGGCACACT	19541
241	GAGCGAGCGGGCCGCGCTAGCGCGAGCGGGTCGGGGAGGCGCGCGGGCGGGCTG	300
19540	GAGCGAGCGGGCCGCGCTAGCGCGAGCGGGTCGGGGAGGCGCGCGGGCGGGCTG	19481
301	ACGTAACCTGCGCGCGCGGAGCTCAGGGCGGGCGGGCGGGCGGGCGGGCGGGCTG	360
19480	ACGTAACCTGCGCGCGCGGAGCTCAGGGCGGGCGGGCGGGCGGGCGGGCGGGCTG	19421
361	CGAACACGCTGGGCACTCCATTCCGGGGCTGTTTACTCCCAACTCTCGCGAGACTGGGCG	420
19420	CGAACACGCTGGGCACTCCATTCCGGGGCTGTTTACTCCCAACTCTCGCGAGACTGGGCG	19361
421	GCGGGCCAGAGGCGCCACAGCTCGGAGCTCAGCTCCGCGCAGCCACAGCTGCCTGTC	480
19360	ACCGGGCCAGAGGCGCCACAGCTCGGAGCTCAGCTCCGCGCAGCCACAGCTGCCTGTC	19301
481	TGTCCCGCGCTCCCGGGGCTTCCGCTCGCGCTCTGGACGCGCTGGGGCAGCGGACACCG	540
19300	TGTCCCGCGCTCCCGGGGCTTCCGCTCGCGCTCTGGACGCGCTGGGGCAGCGGACACCG	19241
541	CGGGAGGATGACGAAGTGTTCGCGACATTTGCGCGCGCGCGGGCGGGCGGGCGGGTGG	600
19240	CGGGAGGATGACGAAGTGTTCGCGACATTTGCGCGCGCGGGCGGGCGGGCGGGTGG	19182
601	AAGCGAGGGCGGTGGCCAGAGCTGCGAGCGCGGAGAACTGCGGGGAAACCTTTG	660
19181	AAGCGAGGGCGGTGGCCAGAGCTGCGAGCGCGGAGAACTGCGGGGAAACCTTTG	19122
661	GTCCGCTCTGCGCGTCTTAGGATCCCGGAAAGAGCAGCGGGCGCGGAAAGCGGCCAG	720
19121	GTCCGCTCTGCGCGTCTTAGGATCCCGGAAAGAGCAGCGGGCGCGGAAAGCGGCCAG	19062
721	GCTGGCGCAGGATCTAGAAAGACTCGCTGGCGCAGGCTCCCTGCCCGCGGGCTGCTG	780
19061	GCTGGCGCAGGATCTAGAAAGACTCGCTGGCGCAGGCTCCCTGCCCGCGGGCTGCTG	19002
781	TCATGAGCTCGTGGAGAGCTCGCTTCCCGCGCGGACCTTCTCTGAGGGGTCCACGTCCA	840
19001	TCATGAGCTCGTGGAGAGCTCGCTTCCCGCGCGGACCTTCTCTGAGGGGTCCACGTCCA	18942
841	GGCAGCGGGCTCGGACACCCACCCCGCGCGGGCAGCTGCCTGGGTGCGCCCTTAAC	900
18941	GGCAGCGGGCTCGGACACCCACCCCGCGCGGGCAGCTGCCTGGGTGCGCCCTTAAC	18882
901	CCGGCGGTAGCTCGTTAAGATGGCGAAGTGTCCGTCCGGAACACCGCAAAACCCCAAT	960
18881	CCGGCGGTAGCTCGTTAAGATGGCGAAGTGTCCGTCCGGAACACCGCAAAACCCCAAT	18822
961	CCCGCTGCGCGACCTCTGACCCCGCGCCCGCAGCAGACAGACTGGGCTCCCGACG	1020
18821	CCCGCTGCGCGACCTCTGACCCCGCGCCCGCAGCAGACAGACTGGGCTCCCGACG	18762
1021	CGCAGCGCTGCGGGGACACCGGTGCGGAAACGAGGAGCCTTTGTAAACCGCACGCTG	1080
18761	CGCAGCGCTGCGGGGACACCGGTGCGGAAACGAGGAGCCTTTGTAAACCGCACGCTG	18702
1081	TTTGCTCTTTTGAAGAAACAGAAATGTTTAAATGTCTGTAAGAGCTTGGCGCCT	1140
18701	TTTGCTCTTTTGAAGAAACAGAAATGTTTAAATGTCTGTAAGAGCTTGGCGCCT	18642
1141	AAAGATGCTCGGTGACTTAGATGTAGGATCAGTTTGTCTGTAAGAGCTTGGCGCCT	1200
18641	AAAGATGCTCGGTGACTTAGATGTAGGATCAGTTTGTCTGTAAGAGCTTGGCGCCT	18582
1201	CCCGGACTCCGTAACCGCACTAGCAGGGGACTGAAAGCGCTTTCAGGTACTGCTGGTGGC	1260

```
Db 18581 CCGGACTCCGTACGSCACTAGCAGGGGACTGAAGCGTCTTCAGGTACTGCTGGTGGGC 18522
Qy 1261 GGTGATGCGCTACAGGCCGATCAGACAGATTTGTGTCTCTGGAACCTTGACACTGACCA 1320
Db 18521 GGTGATGCGCTACAGGCCGATCAGACAGATTTGTGTCTCTGGAACCTTGACACTGACCA 18462
Qy 1321 CGGTAATGCTGAACCTGACCAATATATACAGATCAGACAGCGCATCTCTCTCAACATGA 1380
Db 18461 CGGTAATGCTGAACCTGACCAATATATACAGATCAGACAGCGCATCTCTCTCAACATGA 18402
Qy 1381 TTAAACACAGTTGACTTAATATGTTGGATAAATGTAGAAATCAAAATTACCATACCCAC 1440
Db 18401 TTAAACACAGTTGACTTAATATGTTGGATAAATGTAGAAATCAAAATTACCATACCCAC 18342
Qy 1441 CTCAGGCTCTACTTCGTAAATTTTGAAGGTTGTTTAACTTCTTTGTACCTCAGCTTCT 1500
Db 18341 CTCAGGCTCTACTTCGTAAATTTTGAAGGTTGTTTAACTTCTTTGTACCTCAGCTTCT 18282
Qy 1501 TCATTAACAAAATAGGGGTACTAGCAGCGGGGTGGCTCGGCCCTGTAAATCCAGCACT 1560
Db 18281 TCATTAACAAAATAGGGGTACTAGCAGCGGGGTGGCTCGGCCCTGTAAATCCAGCACT 18222
Qy 1561 TGGGAGGCGGAGCGGACCGGATCACTTCAGGTGAGAAGTTTCAGACCGCCTGTGTCAAC 1620
Db 18221 TGGGAGGCGGAGCGGACCGGATCACTTCAGGTGAGAAGTTTCAGACCGCCTGTGTCAAC 18162
Qy 1621 ATGGGTGAACCGCGCTCTACAAAATATATAAACTTATAGTGTAGTGTGGTAGCGCATG 1680
Db 18161 ATGGGTGAACCGCGCTCTACAAAATATATAAACTTATAGTGTAGTGTGGTAGCGCATG 18102
Qy 1681 ACTGTAATCCAGCAACTCAGGAGGCTGAGGAGAGAAATCGCTTGAACCTGGAGGCGGA 1740
Db 18101 ACTGTAATCCAGCAACTCAGGAGGCTGAGGAGAGAAATCGCTTGAACCTGGAGGCGGA 18042
Qy 1741 GGTTCAGTGTAGTGTAGTGTACCTACCTGCACTCCAGCTTGGCGGACAGAGCGAGACTCT 1800
Db 18041 GGTTCAGTGTAGTGTAGTGTACCTACCTGCACTCCAGCTTGGCGGACAGAGCGAGACTCT 17982
Qy 1801 GCCTTAAATAAATAAATAATTTTAAATAATAGGGGTACTAATATCTACCTTAAAG 1860
Db 17981 GCCTTAAATAAATAAATAATTTTAAATAATAGGGGTACTAATATCTACCTTAAAG 17922
Qy 1861 GATGAGGTTAAATTAAGTACACATAAGCCCTAGCGAGTGGCTTATGSCCTGTAACTCT 1920
Db 17921 GATGAGGTTAAATTAAGTACACATAAGCCCTAGCGAGTGGCTTATGSCCTGTAACTCT 17862
Qy 1921 CAACACTTTGGAGTCTGTGGCGGAGGATCACTTCAGCCAGGAGTTTGAGACTAGTCT 1980
Db 17861 CAACACTTTGGAGTCTGTGGCGGAGGATCACTTCAGCCAGGAGTTTGAGACTAGTCT 17802
Qy 1981 GGGCAACAGAGACATGTCTCTATAGTTGTGTTTGGTTTGTGTTTAACTAGGTTGTGTGT 2040
Db 17801 GGGCAACAGAGACATGTCTCTATAGTTGTGTTTGGTTTGTGTTTAACTAGGTTGTGTGT 17742
Qy 2041 GTGCACCTGCAGTCCAGCTACTAGGGAGGCTGAGGTGGGAGACTGCCTGAGCCACGA 2100
Db 17741 GTGCACCTGCAGTCCAGCTACTAGGGAGGCTGAGGTGGGAGACTGCCTGAGCCACGA 17682
Qy 2101 GGTTCAGGCTGCAGTGAGCCATGATTTGTGCACTGCACTCCAGCTGGGCAACACAGCAA 2160
Db 17681 GGTTCAGGCTGCAGTGAGCCATGATTTGTGCACTGCACTCCAGCTGGGCAACACAGCAA 17622
Qy 2161 GACCTTGTCTCAAAAACAAACAAAGGATCACTCATAAAGTGTCTGGCTCTCTATATGAT 2220
Db 17621 GACCTTGTCTCAAAAACAAACAAAGGATCACTCATAAAGTGTCTGGCTCTCTATATGAT 17562
Qy 2221 CAATATGTGTGTGATCTCTGAATCTCTTCTGACTCAGATCTCATAGATTTTCTGA 2280
Db 17561 CAATATGTGTGTGATCTCTGAATCTCTTCTGACTCAGATCTCATAGATTTTCTGA 17502
Qy 2281 ACTTTTGGAGAACTCTTGGCTCTCTGCAATTTGCAATTCGATCAAGGCACTCCCTCTGCC 2340
Db 17501 ACTTTTGGAGAACTCTTGGCTCTCTGCAATTTGCAATTCGATCAAGGCACTCCCTCTGCC 17442
```

```
Qy 2341 ACCACACAAGCATTTGATTTTAACTTGAATGCTCTCTGTTCCAACTTTAGTAAA 2400
Db 17441 ACCACACAAGCATTTGATTTTAACTTGAATGCTCTCTGTTCCAACTTTAGTAAA 17382
Qy 2401 TTAATCTTGGTCAGGGTCTCTGAAACAGCCCTTTAGTCACTATGCAATTGAATACATGC 2460
Db 17381 TTAATCTTGGTCAGGGTCTCTGAAACAGCCCTTTAGTCACTATGCAATTGAATACATGC 17322
Qy 2461 CCTACAGCT 2469
Db 17321 CCTACAGCT 17313

RESULT 4
AR527527
LOCUS
DEFINITION
Sequence 1 from patent US 6723553.
AR527527
ACCESSION
AR527527.1 GI:53914632
VERSION
KEYWORDS
SOURCE
ORGANISM
Unknown.
Unclassified.
REFERENCE
1 (bases 1 to 980)
AUTHORS
Abrahamsen,L., Ekblom,J., Forsgren,M., Horling,J. and Johansson,P.
TITLES
Promoter sequences
JOURNAL
Patent: US 6723553-A 1 20-APR-2004;
FEATURES
Location/Qualifiers
1..980
source
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN

Query Match 39.7%; Score 980; DB 6; Length 980;
Best Local Similarity 100.0%; Pred. No.1e-199;
Matches 980; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 353 CTCGCGGGGAACACGCTGGGCACTCCATTCGGGGCTGTTTACTCCCAACTCTCGCGAG 412
Db 1 CTCGCGGGGAACACGCTGGGCACTCCATTCGGGGCTGTTTACTCCCAACTCTCGCGAG 60
Qy 413 ACTGCGGCGCGCGGCGAGGCGCCACAGCTGGGAGCCTCAGCTCCGCGCCGACCCAGCGT 472
Db 61 ACTGCGGCGCGCGGCGAGGCGCCACAGCTGGGAGCCTCAGCTCCGCGCCGACCCAGCGT 120
Qy 473 GCCCTGTCTGTCCCGCTCCCGGGCTTTCGCGCGCTCTGGAAGCGCTGGGCGAGCGG 532
Db 121 GCCCTGTCTGTCCCGCTCCCGGGCTTTCGCGCGCTCTGGAAGCGCTGGGCGAGCGG 180
Qy 533 GACACGCGCGGAGGATGGAACGAGTGTCTCGGACATTTTGGCGGCGGGGGCGGGTGG 592
Db 181 GACACGCGCGGAGGATGGAACGAGTGTCTCGGACATTTTGGCGGCGGGGGCGGGTGG 240
Qy 593 CAGGGTGAAGCGGAGGGGCGTGGCAGCGAGCTGCGAGCGCGGCGAGAAACGCGCTGGGG 652
Db 241 CAGGGTGAAGCGGAGGGGCGTGGCAGCGAGCTGCGAGCGCGGCGGAGAAACGCGCTGGGG 300
Qy 653 AACCTTGTCTGTCCGCTCTGCGCGTTCGCTCTAGGATCCCGAAAGGAGACGCGGCGGAAA 712
Db 301 AACCTTGTCTGTCCGCTCTGCGCGTTCGCTCTAGGATCCCGAAAGGAGACGCGGCGGAAA 360
Qy 713 GCGGCGAGCTGGGCCAGGATCTAGAAAGACTGCTGGCGGAGGCTCCCTGCCCGCGG 772
Db 361 GCGGCGAGCTGGGCCAGGATCTAGAAAGACTGCTGGCGGAGGCTCCCTGCCCGCGG 420
Qy 773 GCTGCTGTCTAGTCTGTGGAGAGCTCGCTTCCGCGCGGACCTCTTCTGCAAGGGTTC 832
Db 421 GCTGCTGTCTAGTCTGTGGAGAGCTCGCTTCCGCGCGGACCTCTTCTGCAAGGGTTC 480
Qy 833 CAGCTCCAGGCAACCGCGGCTCGGACACCCCAACCCCGCGGCGGCACTGCGCTGGGTGC 892
Db 481 CAGCTCCAGGCAACCGCGGCTCGGACACCCCAACCCCGCGGCGGCACTGCGCTGGGTGC 540
```


QY 1044 GTGCGTGGCAAGCGGAGGACCTTTGTAACGCCACGTTGTTGCTCTTTTGAAGAAACAAG 1103
 Db 421 GTGCGTGGCAAGCGGAGGACCTTTGTAACGCCACGTTGTTGCTCTTTTGAAGAAACAAG 480
 QY 1104 AATAAATGTTAAATGTTCTGTAAGAAAGCTTGGCGCTTAAAGA 1146
 Db 481 AATAAATGTTAAATGTTCTGTAAGAAAGCTTGGCGCTTAAAGA 523

RESULT 6

HS358H7/c 148864 bp DNA linear PRI 12-DEC-2002
 LOCUS Human DNA sequence from clone Rp3-358H7 on chromosome X, complete
 DEFINITION sequence.

ACCESSION 277249

VERSION 277249.2 GI:26800831

KEYWORDS HTG.

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 148864)

REFERENCE

AUTHORS Pavitt,R.

TITLE Direct Submission

JOURNAL Submitted (05-DEC-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 On Dec 14, 2002 this sequence version replaced gi:1460066.

COMMENT

----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Web site: http://www.sanger.ac.uk
 Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest, except on the rare
 occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information
 on the WORMPEP database can be found at
 http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome X, constructed by the Sanger Centre Chromosome X Mapping
 Group. Further information can be found at

http://www.sanger.ac.uk/HGP/ChrX
 RP3-358H7 is from the library RPCI-3 constructed by the group of
 Pieter de Jong. For further details see
 http://www.choxi.org/bacpac/home.htm
 VECTOR: pCVPAC2.

FEATURES

source

1. 148864
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="RZPD:RPCIP704H07358"
 /db_xref="taxon:9606"
 /chromosome="X"
 /clone="Rp3-358H7"
 /clone_lib="RPCI-3"

ORIGIN

Query Match 14.1%; Score 347; DB 9; Length 148864;

Best Local Similarity 70.7%; Pred. No. 1.2e-63;
 Matches 503; Conservative 0; Mismatches 205; Indels 3; Gaps 3;

QY 1509 AAAATAGGGGTAATAAGCAGCGGGGGTCTCGCCCTGTGTAATCCAGACACTTGGGAGG 1568

Db 77689 AAAATGTTGGGGTGGCCAGGTGTGGTCTCACACCTGTGTAATCCCAACACTTTGGGAGG 77630

QY 1569 CCGAGGCGCGGATCACTTGGAGTTCAGAGTTTCAGACCGAGCTGGTCAACATGGGTCA 1628

Db 77629 CCGAGGCGGGTGGATCACTTGGAGTTCAGAGTTTGGGACCGACCTGTCCAACAT-GGTGA 77571

QY 1629 AACCGCGGCTTACCAAAAAATAAAAACTTAGCTGAGTGTGGTAGCGCATGACTGTGTAAT 1688

Db 77570 AACCCGGTCTTACTTAAAAACACAAAAATTAGCCAGGGTGGCAGCAGGTGCCAGGAAT 77511

QY 1689 CCCAGCAACTCAGGAGGCTGAGGCA-GAGAAATCGTGTGAACCTGGGAGCGGAGGTTGCA 1747

Db 77510 CCCAGCTACTCGGAGGCTGAGGAGGAGAAATCGCTTGAACCCGGGAGACAGAGGTTGCA 77451

QY 1748 GTGAGCTGAGATCGTACCACTGCACCTCCAGCTTGGGCGAC-AGAGCGAGACTCTGCCTTA 1806

Db 77450 GTGAGCCAAGATCGTGCCACTGCACCTCCAGCTTGGGTGACGAGAGGAAATCCCGTCTCA 77391

QY 1807 AAAATAAATAAATAAATTTTTTAAAAAATAAGGGGTACTAATAATCTACCTTAAAGGATGAG 1866

Db 77390 AAAAAAATAAATAAATAAATAAGTATGGGTGATGTCTTCAAGTGGCATCTTACCCGTGGAAT 77331

QY 1867 GGTAAATTAAGTACACATTAAGCCCTAGCCAGTGGCTTATGCTGTAAATCTCAACAC 1926

Db 77330 TGCCTCTTTAGGAGGATGGTAGAGGCCAGGCACAGTGGTCAACCTGTTCATCCCAACAC 77271

QY 1927 TTTGGAGTCTGTGGGGGAGGATCACTTGGCCAGGAGTTTGAGACTAGTCTGGGCAA 1986

Db 77270 TTTGGGATGCTGGGGTGGGAGGATTCCTTGGCCAGGAAATTCAGACTAGTCTGGGCAA 77211

QY 1987 CAGAGACATGCTCTATAGTGTGTTGGTTGTTGTTTATCACCAGGTGTGGTGTGTCAC 2046

Db 77210 GATAGTGACACCCCTCTATACAAATAGTTAAATAATCAGTCAGCATGGCGTGCACAC 77151

QY 2047 CTGCACTCCAGCTACTAGGAGGCTGAGGTGGGAGGACTGCTGAGCCAGAGGTGCA 2106

Db 77150 CTGTAGTCTCTCTACTTGGGAGGCTGAGGTGGGAGGATCACATGATGTCATGAGTTGA 77091

QY 2107 GCCTCAGTGGCCATGATTGTCACCTGCACCTCCAGCTGGGCAACAGCAAGACCTT 2166

Db 77090 GGTCTCAGTGGGATGATCGTGTCACTGCATCTCAGCTGGGGCAACAGCAAGACCTT 77031

QY 2167 GTCTCAAAAAACAAACAAAGCATACTCATAAAGTGTCTCGCTCTCTATATG 2217

Db 77030 GTCTCAAAAAACAAACAAAGAAAGAAAGAAAGAGGAGGATGGTAAAGG 76980

RESULT 7

AC010519

LOCUS Homo sapiens chromosome 19 clone LLNL-261D9, complete sequence.

DEFINITION 46741 bp DNA linear PRI 14-JUL-2002

ACCESSION AC010519

VERSION AC010519.8 GI:21747432

KEYWORDS HTG.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 46741)

REFERENCE

AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.

TITLE Direct Submission

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 46741)

AUTHORS DOE Joint Genome Institute.

TITLE Direct Submission

JOURNAL Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE 3 (bases 1 to 46741)

AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
 TITLE Direct Submission
 JOURNAL Submitted (02-JUN-2000) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE 4 (bases 1 to 46741)
 TITLE DOE Joint Genome Institute and Stanford Human Genome Center.
 JOURNAL Direct Submission
 Submitted (17-OCT-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE 5 (bases 1 to 46741)
 TITLE DOE Joint Genome Institute and Stanford Human Genome Center.
 JOURNAL Direct Submission
 Submitted (14-JUL-2002) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

COMMENT On Jul 14, 2002 this sequence version replaced gi:16195205.
 Draft Sequence Produced by DOE Joint Genome Institute
 www.jgi.doe.gov
 Finishing Completed at Stanford Human Genome Center
 www.shgc.stanford.edu
 Quality: Phrap Quality >=40 99.8% of Sequence;
 Estimated Total Number of Errors is 0.1.

FEATURES
 source
 1. .46741
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="19"
 /clone="L1NLR-261D9"

ORIGIN
 Query Match 14.0%; Score 346; DB 9; Length 46741;
 Best Local Similarity 71.4%; Pred. No. 1.7e-63;
 Matches 510; Conservative 0; Mismatches 200; Indels 4; Gaps 4;

QY 1480 CCTCTTTGTACCTCAGCTTCTTCATTACAAAATAGGGTACTAGCCAGCGGGTGGCT 1539
 DB 40713 CTTCTTTGTAGCCTCATCTGCAAAATGTGGATCAATCACCTTGGGCGAGCGCGTGGCT 40772

QY 1540 CGGCGCTGTATCCAGCACTTGGGAGCGCGGAGCGGATCACTTGGGTCTAGAG 1599
 DB 40773 CACACCTGTATCCAGCGCTTTGGGAGCGCAAGCGAGGTAGTACTTGGGCCAGGAG 40832

QY 1600 TTTCCAGACAGCTGGTCAACATGGGTGAACCGCGCTCTACCAAAATATAAAACATT 1659
 DB 40833 TTGGAGACAGCTGGCCAACT-GGTGAACCGCTCTCTACCAAAACACAAAATTT 40891

QY 1660 AGCTGAGTGTGGTAGCGCACTACTGTAAATCCAGCACTCAGAGGCTGAGGAG-AGAA 1718
 DB 40892 AGCTGGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGT 1778

QY 1719 TCGTTGAACCTGGGAGCGGAGGTTCAGTGAGCTGAGATCGTACCAGCTGCATCCAGC 1778
 DB 40952 TTGCTTAAACCTGGGAGCGGAGGTTCAGTAAAGCGAGACCGTGGCCACTGCATTCGAGC 41011

QY 1779 TTGGGCGACAGCGAGCACTCTGCTTTAAATAATAATAATAATAATAATAATAATAAGG 1838
 DB 41012 CTGGGCAACAGAGCGAGCACTGTCTTAAATAATAATAATAATAATAATAATAAGGATCAGATCA 41071

QY 1839 GGTACTAATATCTACCTTTAAAGGATGAGGGTTAAATTAAGTACACATAGCCCTAGCG 1898
 DB 41072 CCATGGTCTATGCCATGGTGTGCTGAGGACTTGAATAGCAC-CTGGTGGCGGGCG 41130

QY 1899 CAGTGGCTTATGCTGTAACTCAACACTTTGGGAGTCTGTGGCGGAGATCACTTTAG 1958
 DB 41131 CAGTGTCTCAGCTGTAAATCCAGCACTTTGGGAGGCTTGGGCGGCGAGATCACTTTAG 41190

QY 1959 CCAGGAGTGTGAGCACTGTCTGGGCAACAGACATGTCTCTATAGTGTGTTGTTT 2018
 DB 41191 GCCAGAGTGTGAGCAGCAGCTGGCCCAACATGTGTGAACCCCGTCTCTACTAAATAACA 41250

QY 2019 TGTATTTACAGGTGTGGTGTGTGACCTGCGAGTCCAGCTACTAGGAGGCTGAGGTG 2078
 DB 41251 AAATTAGCTGGGTGAGTGGTGGCCACCTGTAAATCCAGCTACTTCGGGAGGCTTAGGCA 41310

QY 2079 GGAGGACTGCTGAGCCCGAGGAGTTCGAGGCTGCGAGTGCAGCCATGATTTGTCCACTGCAC 2138
 DB 41311 GGAGGATCATTGAATCGGAGGCGGAGGTTGCGAGTGCAGTGCAGATAGTGCATTGCAC 41370

QY 2139 TCCAGCCTGGGCAAC-ACAGCAAGACCTTGTCTCAAAACAAACAAACAAACGATA 2191
 DB 41371 TCCAGCCTGGGCGCAACAGCAAACTCTGTCTCAAAACAAACAAACAAACAAAAA 41424

RESULT 8
 AC093324/c
 LOCUS
 DEFINITION Homo sapiens chromosome RPCI-11 clone RP11-565M3, WORKING DRAFT
 SEQUENCE, 43 unordered pieces.
 AC093324
 AC093324.1 GI:15209236
 HTG: HTGS_PHASE1; HTGS_DRAFT.
 KEYWORDS Homo sapiens (human)
 SOURCE Homo sapiens
 ORGANISM
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 194241)
 Waterston,R.H.
 The sequence of Homo sapiens clone
 Unpublished
 2 (bases 1 to 194241)
 Waterston,R.H.
 Direct Submission
 Submitted (18-AUG-2001) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA

COMMENT
 ----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc/index.shtml
 ----- Project Information -----
 Center project name: H_NH0565M03
 ----- Summary Statistics -----
 Sequencing vector: M13; 0%
 Sequencing method: plasmid; 100%
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 171387 bases at least Q40
 Consensus quality: 176465 bases at least Q30
 Consensus quality: 179780 bases at least Q20
 Insert size: 210000; agarose-fp
 Quality coverage: 3.09 in Q20 bases; agarose-fp
 Quality coverage: 3.56 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 43 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 1215: contig of 1215 bp in length
 * 1216 1315: gap of unknown length
 * 1316 3125: contig of 1810 bp in length
 * 3126 3225: gap of unknown length
 * 3226 4721: contig of 1496 bp in length
 * 4722 4821: gap of unknown length
 * 4822 6056: contig of 1235 bp in length
 * 6057 6156: gap of unknown length
 * 6157 7899: contig of 1743 bp in length
 * 7900 9391: gap of unknown length
 * 8000 9391: contig of 1392 bp in length

Db	64338	CTTCTTGTAGCTCATCTGCAAAATATGTGGATCAATCACCTTGGGCCAGGCCGGTGGCT	64279
Qy	1540	CGCGCTGTAAATCCAGCACTTGGGAGGCGGAGGCGGATCACTTGGAGTCAAGAG	1599
Db	64278	CACACTGTAAATCCAGCGCTTGGGAGGCGGAGGCGGATCACTTGGAGGCGGAGGAG	64219
Qy	1600	TTTCAGACAGCGCTGTCAACATGGGTGAAACGCCGGCTCTACCAAAAAATATAAAACTT	1659
Db	64218	TTGGAGACAGCGCTGGCCACAT-GGTGAAACCCCGTCTCTACCAAAAAACACAAAAAAT	64160
Qy	1660	AGCTGAGTGTGTAGCGCATGACTGTAATCCAGCACTCAGAGAGCTGAGGAG-AGAA	1718
Db	64159	AGCTGGGT	64100
Qy	1719	TCCTTTGAACCTGGGAGGCGGAGTTCAGTGAAGTGCAGTGCAGTGCAGTGCAGTGCAG	1778
Db	64099	TTGCTTAAACCTGGGAGGCGGAGTTCAGTGAAGTGCAGTGCAGTGCAGTGCAGTGCAG	64040
Qy	1779	TTGGGCGCACAGAGCGAGCTCTGCCTTAAATAAATAAATAAATAAATAAATAAATAA	1838
Db	64039	CTGGGCAACAGAGCGAGCTCTGCTTAAATAAATAAATAAATAAATAAATAAATAA	63980
Qy	1839	GGTACTAATATCTACCTTAAGGATGAGGTAAATTAAGTACACATGAAGCCCTAGCG	1898
Db	63979	CATGGTCTCTATGCCATGGTGTGTGTGAGG--ACTTGGAAATAGCACCTGTGTGGCGGG	63922
Qy	1899	CAGTGGCTTATGCTGTAAATCTCAACACTTTGGGAGTCTGTGGCGGAGGATCACTTGAG	1958
Db	63921	CAGTGTCTCAGCGCTGTAAATCCAGCACTTTGGGAGGCTGAGGCGGCGAGATCACTTGAG	63862
Qy	1959	CCAGGAGTTTGAGACTAGTCTGGGCAACAGAGACATGCTCTATAGTTGTGTGTGTGT	2018
Db	63861	GCAGAGTTTGAGACAGCGCTGGCAACATGTGTGAACCCGCTCTACTAAATAACA	63802
Qy	2019	TGTTTTTACCAGTGTGGTGTGACCTTGGAGTCTGAGTCCAGTCTAGGAGGCTGAGGTG	2078
Db	63801	AAAATTAGCTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG	63742
Qy	2079	GGAGGACTGCTGAGCCAGAGTGTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGC	2138
Db	63741	GGAGATCACTTGAATCGGGAGGCGGAGGTTCAGTGTGAGTGTGAGTGTGAGTGTGAG	63682
Qy	2139	TCAGGCTGGGCAAC-ACAGCAAGCACTTGTCTCAAAAAACAAACAAAAACATA	2191
Db	63681	TCAGGCTGGGCAACAGCAAGCACTTGTCTCAAAAAACAAACAAAAACAAAAA	63628
RESULT 9			
AC108000		167587 bp	DNA linear PRI 21-MAY-2002
DEFINITION			Homo sapiens chromosome 15, clone CTD-2116G1, complete sequence.
AC108000			
VERSION		AC108000.5	GI:21039854
KEYWORDS			HTG.
SOURCE			Homo sapiens (human)
ORGANISM			Homo sapiens
REFERENCE			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS			1 (bases 1 to 167587)
TITLE			Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
JOURNAL			Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B.,
AUTHORS			Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
			Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
			Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
			Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
			Ginde, S., Gordon, L., Goyette, M., Graham, L., Grand-Pierre, N.,
			Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
			Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R.,
			Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C.,
			Macdonald, P., Macdonald, P., Major, J., Marquis, N.,
			Mathews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J.,
			Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
			Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P.,
			O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
			Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C.,
			Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S.,
			Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N.,
			Stojanovic, N., Straus, N., Subramanian, A., Talamas, J., Testaye, S.,
			Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J.,
			Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
			Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
			Direct Submission
			Submitted (21-MAY-2002) Whitehead Institute/MIT Center for Genome
			Research, 320 Charles Street, Cambridge, MA 02141, USA
			3 (bases 1 to 167587)
			Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
			Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L.,
			Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
			Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
			Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
			Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
			Ginde, S., Gordon, L., Goyette, M., Graham, L., Grand-Pierre, N.,
			Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
			Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R.,
			Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C.,
			Macdonald, P., Macdonald, P., Major, J., Marquis, N.,
			Mathews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J.,
			Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
			Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P.,
			O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
			Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C.,
			Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S.,
			Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N.,
			Stojanovic, N., Straus, N., Subramanian, A., Talamas, J., Testaye, S.,
			Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J.,
			Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
			Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
			Direct Submission
			Submitted (28-APR-2002) Whitehead Institute/MIT Center for Genome
			Research, 320 Charles Street, Cambridge, MA 02141, USA
			4 (bases 1 to 167587)
			Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
			Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L.,
			Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
			Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
			Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
			Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
			Ginde, S., Gordon, L., Goyette, M., Graham, L., Grand-Pierre, N.,
			Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
			Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R.,
			Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C.,
			Macdonald, P., Macdonald, P., Major, J., Marquis, N.,
			Mathews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J.,
			Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
			Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P.,
			O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
			Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C.,
			Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S.,
			Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N.,
			Stojanovic, N., Straus, N., Subramanian, A., Talamas, J., Testaye, S.,
			Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J.,
			Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
			Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
			Direct Submission
			Submitted (21-MAY-2002) Whitehead Institute/MIT Center for Genome
			Research, 320 Charles Street, Cambridge, MA 02141, USA
			On May 21, 2002 this sequence version replaced gi:20336146.
			All repeats were identified using RepeatMasker:
			Smit, A.P.A. & Green, P. (1996-1997)
			http://ftp.genome.washington.edu/RM/RepeatMasker.html
			----- Genome Center
			Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center Project name: L24580
 Center Clone name: 2116_G_1

FEATURES	Location/Qualifiers
source	1..167587
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="15"
	/map="15"
	/clone="CTD-2116G1"
	/clone_lib="CITDI Human BAC"
	complement(11..144)
repeat_region	/rpt_family="L2"
repeat_region	complement(145..356)
repeat_region	/rpt_family="AluSg/x"
repeat_region	complement(357..1109)
repeat_region	/rpt_family="L2"
repeat_region	3713..3828
repeat_region	/rpt_family="L1ME"
repeat_region	complement(3834..4006)
repeat_region	/rpt_family="MIR"
repeat_region	4574..4603
repeat_region	/rpt_family="(GAAA)n"
repeat_region	complement(4976..5167)
repeat_region	/rpt_family="MIR"
repeat_region	5639..5775
repeat_region	/rpt_family="L1"
repeat_region	6745..6766
repeat_region	/rpt_family="AT_rich"
repeat_region	complement(8290..8370)
repeat_region	/rpt_family="MIR3"
repeat_region	9077..9288
repeat_region	/rpt_family="MIR"
repeat_region	9496..10838
repeat_region	/rpt_family="L1MD1"
repeat_region	10839..11162
repeat_region	/rpt_family="AluSx"
repeat_region	11163..12229
repeat_region	/rpt_family="L1MD1"
repeat_region	complement(13118..13414)
repeat_region	/rpt_family="AluY"
repeat_region	13667..14061
repeat_region	/rpt_family="L2"
repeat_region	14366..14395
repeat_region	/rpt_family="AT_rich"
repeat_region	14748..15052
repeat_region	/rpt_family="AluSx"
repeat_region	complement(15603..15768)
repeat_region	/rpt_family="MIR"
repeat_region	complement(15878..15932)
repeat_region	/rpt_family="L2"
repeat_region	complement(15988..16069)
repeat_region	/rpt_family="MER81"
repeat_region	complement(16309..16442)
repeat_region	/rpt_family="MER46A"
repeat_region	complement(16557..16730)
repeat_region	/rpt_family="MIR"
repeat_region	complement(16746..16830)
repeat_region	/rpt_family="MIR"
repeat_region	16845..16939
repeat_region	/rpt_family="MER117"
repeat_region	complement(17267..17436)
repeat_region	/rpt_family="MIR3"
repeat_region	complement(17914..18014)
repeat_region	/rpt_family="MER103"
repeat_region	complement(19124..19415)
repeat_region	/rpt_family="AluJo"
repeat_region	complement(19571..19636)

repeat_region	/rpt_family="L2"	19637..19926
repeat_region	/rpt_family="AluJo"	complement(19927..19984)
repeat_region	/rpt_family="L2"	complement(20449..21788)
repeat_region	/rpt_family="L2"	21985..22144
repeat_region	/rpt_family="MIR"	22728..22838
repeat_region	/rpt_family="Charlie8"	23014..23044
repeat_region	/rpt_family="AT_rich"	23140..23256
repeat_region	/rpt_family="FLAM_C"	23265..23350
repeat_region	/rpt_family="(TA)n"	23357..23381
repeat_region	/rpt_family="AT_rich"	24908..25206
repeat_region	/rpt_family="AluSx"	complement(25805..25979)
repeat_region	/rpt_family="L1MB4A"	complement(26043..26262)
repeat_region	/rpt_family="L1MB4A"	complement(26861..26993)
repeat_region	/rpt_family="FLAM_C"	complement(27023..27317)
repeat_region	/rpt_family="AluSx"	27952..28300

Query Match 13.9%; Score 343.4; DB 9; Length 167587;
 Best Local Similarity 75.3%; Pred. No. 6.9e-63;
 Matches 507; Conservative 0; Mismatches 156; Indels 10; Gaps 6;

QY	1519	TACTAGCCAGGCGGGTGGCTCGCGCTGTATATCCAGCAGCTTGGGAGGCGGAGCAGC	1578
Db	83398	TACTGCCAGGCGCAGTGGCTCAGCGCTGTATATCCAGCAGCTTGTGAGGCGGAGGTGG	83457
QY	1579	CGATCAGCTTGAGTGCAGAGTTTCAGACCAGCTGGTCAACATGGGTGAACCGCGCT	1638
Db	83458	TGGATCAGCTTGAGTGCAGAGTTTCAGACCAGCTGGTCAACATGGGTGAACCGCGCT	83516
QY	1639	CTACCAAAATATAAACTTAGCTGAGTGGTGGAGCAGTCACTTAATCCAGCAACT	1698
Db	83517	CTACTAAAAATACAAAA-TTAGCTGGGTGGTGGCGGTGACTGTATCCAGCACT	83575
QY	1699	CAGGAGGCTGAGGCA-GAGATCGCTTGAACCTGGAGGCGGAGGTTCAGTGCAGT	1757
Db	83576	CAGGAGGCTGAGGAGGAGATCACTTGAACCGGAGGCGGAGGTTCAGTGCAGTGCAG	83635
QY	1758	ATCGTACCACTGCAGCTCCAGCTTGGGCGACAGCGAGACTCTGCCTTAAAAATAAATA	1817
Db	83636	ATTGCACCACTGCAGCTCCAGCTTGGGCGACAGCGAGACTCTGCCTTAAAAATAAATA	83695
QY	1818	ATAATTTTAAAAAATAGGGTACTAATATCTACCTTAAAGGATGAGGGTTAAATTA	1877
Db	83696	ATAATAATAAATAAATAAGACCTCC--CTTAAAGGCAAGGTCAAAGATATGAAGAA	83753
QY	1878	GTACACACATAAGCCCTAGCGAGTGGCTTATGCCCTGTATCTCAACACTTTGGGAGTCT	1937
Db	83754	TTCAGATGGCAGGCGGCGCACAGTGGCTTATGCCCTGTATATACCACTTTCTTTGGGAGGC	83813
QY	1938	GTGGCGGAGGATCACTTGAAGCCAGGAGTTGAGACTAGTCTGGGCAACAGAGACATGT	1997
Db	83814	AAGGTGGTGGATCACCTGAGGTTCAGAGTTTATAGACCGCTGGGCAACATGGTGAAC	83873
QY	1998	CTCTATAGTTGTTTGGTTTGTGTTTAA-CAGAGTGTGGTGTGTGACCTGCAGTCCC	2056
Db	83874	CCCTCTCTACTAAAAATATAAAAAATAGCTGGGTGGTGGCGGATGCCTGTATATCC	83933
QY	2057	AGCTACTAGGAGGCTGAGGTGGGAGGACTGCCTGAGCCAGGAGGTGAGGCTGAGTG	2116

Db	30678	AAAAAAAAAAAAAAAAAGAAAAG	30656	
RESULT 11				
AC010605	86684 bp	DNA	linear	PRI 18-APR-2000
Homo sapiens chromosome 19 clone CTFB-18985, complete sequence.				
AC010605	ACCESSION			
AC010605.4	GI:7019286			
HTG.				
Homo sapiens (human)				
Homo sapiens				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
DOE Joint Genome Institute and Stanford Human Genome Center.				
Direct Submission				
Unpublished				
DOE Joint Genome Institute.				
2 (bases 1 to 86684)				
Direct Submission				
Submitted (16-SEP-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA				
3 (bases 1 to 86684)				
DOE Joint Genome Institute and Stanford Human Genome Center.				
Direct Submission				
Submitted (22-FEB-2000) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA				
4 (bases 1 to 86684)				
DOE Joint Genome Institute and Stanford Human Genome Center.				
Direct Submission				
Submitted (18-APR-2000) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA				
On Feb 22, 2000 this sequence version replaced gi:6910541.				
Draft Sequence Produced by DOE Joint Genome Institute				
www.jgi.doe.gov				
Finishing Completed at Stanford Human Genome Center				
www.ehgc.stanford.edu				
Quality: Phrap Quality >=40 99.7% of Sequence;				
Estimated Total Number of Errors is 0.5.				
Location/Qualifiers				
1. 86684				
/organism="Homo sapiens"				
/mol_type="genomic DNA"				
/db_xref="taxon:9606"				
/chromosome="19"				
/clone="CTB-18985"				
FEATURES				
source				
Query Match				
Best Local Similarity				
Matches 496; Conservative 0; Mismatches 176; Indels 11; Gaps 4;				
13.7%; Score 339.4; DB 9; Length 86684;				
72.6%; Pred. No. 4.7e-62;				
Query				
1519 TACTAGCCAGCGCGGTGGCTCGCGCTGTAATCCACGACTTTGGGAGCCGAGCAGC 1578				
Db				
38703 TCCAGGCGGGCATGATGGCTCAGCGCTGTAATCCAGAACTTTGGAGGCTGAGCAGG 38762				
Query				
1579 CGGATCACTTGAGTTCAGAGTTTCAGACAGCGCTGTCACATGGGTGAAACCGCGCT 1638				
Db				
38763 CGGATCACTTGAGTTCAGAGTGTGAGACAGCGCTGGCCAAAAT-GGCGAAATCCGCTCT 38821				
Query				
1639 CTACAAAAATATAAAACCTTAGCTAGTGTGTGTAGCGCATGACTGTAATCCCGAACCT 1698				
Db				
38822 CTACTGAATATACAAAAA-TTAGCCGGGCATGTGGCGCATGCTGTGATCCAGCTACT 38880				
Query				
1699 CAGGAGGCTGAGGCA-GAGAAATCGTTGAAACCTGGGAGCGGAGTTTGCAGTGCAGTGAG 1757				
Db				
38881 CGGGAAGCTGAGGCATGAGAATCACTTCAACAAGGGGAGGAGGTTGCAGTGCAGTGAG 38940				
Query				
1758 ATCGTACCACTGCATCCAGCTTGGGCGCAGACGCGAGTCTGCTTTAAAAATAATATA 1817				
Db				
38941 ACCTCACCACCTGCATCCAGCCTTAGGCGCAGAGAGTGCAGAGCTGTCTTAATAATAATA 39000				
QY	1818	ATAATTTTAAAAAATAGGGGTACTAATATCTACCTTTAAAGATGAGGGTTAAATTTAA	1877	
DB	39001	TAAATAAAACCTCTCTTTTGACCCCTCTTAGAATCTCTCTCCACAAATGCAAGAAATTTAA	39060	
QY	1878	GTACACACATAAGCCCTAGCGAGTGGCTTATGCTGTAATCTCAACACTTTTGGAGTCT	1937	
DB	39061	-----ATGGGTCTCTGGCATGGGGCTCACCTGTAATCCAGCACTTTGGAGGCC	39112	
QY	1938	GTGGCGGAGGAGTCACTTTGAGCCAGGAGTTTGAGACTAGTCTGGGCAACAGACATGT	1997	
DB	39113	GAGGCAGGAGGATCTCTTGAGCCAGGAGTTTGAGACCGCGCTGGCAACACAGTGAAC	39172	
QY	1998	CTCTATAGTTGTGTTGTTGTTTATACAGGTGTGGTGTGTGACCTGCGAGTCCCA	2057	
DB	39173	CCCGTCTCTACAAAAAATTTAAAAAGTAGCTGCAATAGTGTGTGCACTGTAGTCTCA	39232	
QY	2058	GCTACTAGGGAGGCTGAGGTGGGAGGACTGCTTGAGCCAGGAGGTGAGGCTCGAGTGA	2117	
DB	39233	GCTACTTGGGAGGCTGAGGTGGGAGGATGGCTTGAGCCTGGGAGGTTGAGGCTCGAGTGA	39292	
QY	2118	GCCATGATTGTGCCACTGCACCTCCAGCGCTGGGCAACACAGCAAGACCTTGTCTCAAAAAC	2177	
DB	39293	GCTGTGATCATGCCACTGCACCTTCACTTGGCTGGCTGACATGCAGACCTGTCTCAAAAAA	39352	
QY	2178	AAACAAAAAGCATCTCTATAAG	2200	
DB	39353	AAAAAAAAAAAAAAAAAGAAAAG	39375	
RESULT 12				
AC080005	169583 bp	DNA	linear	PRI 03-SEP-2002
Homo sapiens 3 BAC RP11-475N22 (Roswell Park Cancer Institute Human BAC Library) complete sequence.				
AC080005	ACCESSION			
AC080005.24	GI:22024355			
HTG.				
Homo sapiens (human)				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
1 (bases 1 to 169583)				
REFERENCE				
AUTHORS				
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C., Alsbrooks, S.L., Anaratunge, H.C., Are, J.R., Ayale, M., Banks, T., Barbaria, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., He, X., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homs, P., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louissegh, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapu, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhinney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Moore, S., Morgan, M., Moorish, T., Morris, S., Moser, M., Neal, D., Nelson, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokoko, S., Ogih, M., Okuwonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Peery, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M.,				

repeat_region		/function="low quality"	
13623..13639		/rpt_family="(ATGGTG)n"	
repeat_region		13640..14059	
14060..14120		/rpt_family="MSTA"	
repeat_region		14127..14543	
14127..14543		/rpt_family="(ATGGTG)n"	
repeat_region		14127..14543	
14127..14543		/rpt_family="L2"	
repeat_region		complement(14803..15276)	
15281..15439		/rpt_family="MLTIP"	
repeat_region		15281..15439	
Query Match		13.7%; Score 339.4; DB 9; Length 169583;	
Best Local Similarity		59.6%; Pred.No. 5e-62;	
Matches 517; Conservative		0; Mismatches 221; Indels 5; Gaps 4;	
QY	1447	CTTCTACTTCGTAATTTTGAGCAGGTTGTTTAACTCTTTTGATACCTCAGTCTCTCTCATTA	1506
DB	163181	CCTCAATTCCTTTTAAAGATTATTTCTGTGCGCCAGCAGTGCCTCTCTCTTTC	163240
QY	1507	CAAAATAGGGTACTAGCAGCGGGTGGCTGGCCCTGTAAATCCAGCACTTGGGA	1566
DB	163241	AAAAATCAGAAAGCTGCCTGGTGTGTGGCTCACACCTGTAACTCCAGCACTTAGGA	163300
QY	1567	GGCCGAGGCGGATCACTTGAGCTCAGAGTTTTCAGACCACTGCTGCTCAACATGGCT	1626
DB	163301	GGCCGAGGCGGAGATCACTTGAGCTCAGAGTTTTCAGACCACTGCTGCTCAACATGGCT	163359
QY	1627	GAACCCCGGCTTACCAAAATATATAAACTTAGCTGAGTGGTGGTGGCATGACTGTA	1686
DB	163360	AAACCTCATCTTTTACATAAAATACAAAATAGCAGTATGGTACCTGTGCTTATA	163419
QY	1687	ATCCAGCAACTCAGAGGCTGAGGA-GAGAAATCCTTGAACCTGGGAGGCGAGGTTG	1745
DB	163420	ATCCAGCACTCAGGAGGCTGAGGAGGAGAAATGCTTGAACCCAGGAGGAGAGTTTG	163479
QY	1746	CAGTGAAGTGAATCTACACTGCACTCCAGCTTGGGGGACAGAGGAGACTCTGCTT	1805
DB	163480	CAGTGAAGCAAGATTGCACTGCACTCCAGCTTGGGGGACAGAGGAGGA--CTGCTG	163537
QY	1806	AAAAATAATAATAATTTTAAAAAATAGGGTACTAATATCTACCTTAAAGGATGA	1865
DB	163538	AAAAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA	163597
QY	1866	GGTTAAATTAAGTACACATTAAGCCCTAGCCAGTGGCTTATGCTGTATATCTCAACA	1925
DB	163598	AGAAAAACGAAAAAGAAAGCTGGGGGGGACAGTGGCTCATGCTGTATATCCCAACA	163657
QY	1926	CTTTGGGAGTCTGTGGGGAGGATCACTTGAGCCAGGAGTTTGAGACTAGTCTGGGA	1985
DB	163658	CTTTGGGAGGCAAGTGGGAGATACCTTGAGCCAGGAGTTTCGAGAACAGGCTGGCA	163717
QY	1986	ACAGAGACATGCTCTATAGTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTA	2045
DB	163718	GCATAG-CAAAACCTGCTCTACAAAAATACAAAAATAGCAGTGGTGGTGTGTA	163776
QY	2046	CTTGAGTCCAGCTACTAGGAGGCTGAGTGGAGGAGTGGTGGAGGAGGAGGAGGCTG	2105
DB	163777	CCTGTGCTCTAGTACTTCTGAGTCTGAGTGGAGGATGCTTAAAGCTGGGAGGCTG	163836
QY	2106	AGCTGAGTGGAGGATGTTGCCACTGCACTCCAGCTGGGACACAGCAAGACCT	2165
DB	163837	AGTTGCGAGGAGGAGGATGTTGCCACTGCACTCCAGCTGGGACACAGGAGGAGCT	163896
QY	2166	TGCTCAAAAACAAACAAAAGC	2188
DB	163897	TGCTCAAAAACAAAAAATGC	163919
RESULT 13		AC020715/c	
LOCUS		AC020715 197992 bp DNA linear HTG 17-AUG-2000	

DEFINITION		Homo sapiens chromosome 19 clone RP11-446K10, WORKING DRAFT	
ACCESSION		AC020715	
VERSION		AC020715.4 GI:9838201	
KEYWORDS		HTG; HTGS_PHASE1; HTGS_DRAFT.	
SOURCE		Homo sapiens (human)	
ORGANISM		Homo sapiens	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS		Waterston,R.H.	
TITLE		The sequence of Homo sapiens clone	
JOURNAL		Unpublished	
REFERENCE		2 (bases 1 to 197992)	
AUTHORS		Waterston,R.H.	
TITLE		Direct Submission	
JOURNAL		Submitted (08-JAN-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA	
COMMENT		On Aug 17, 2000 this sequence version replaced gi:8570194.	
		----- Genome Center -----	
		Center: Washington University Genome Sequencing Center	
		Center code: WUGSC	
		Web site: http://genome.wustl.edu/gsc/index.shtml	
		----- Project Information -----	
		Center project name: H_NH0446K10	
		----- Summary Statistics -----	
		Sequencing vector: M13; #	
		Sequencing vector: plasmid; #	
		Chemistry: Dye-primer ET; # of reads	
		Chemistry: Dye-terminator Big Dye; # of reads	
		Assembly program: Phrap; version 0.990319	
		Consensus quality: 193859 bases at least Q40	
		Consensus quality: 194948 bases at least Q30	
		Consensus quality: 195923 bases at least Q20	
		Insert size: 209000; agarose-fp	
		Insert size: 197594; sum-of-contigs	
		Quality coverage: 6.86 in Q20 bases; agarose-fp	
		Quality coverage: 7.28 in Q20 bases; sum-of-contigs	
		* NOTE: This is a 'working draft' sequence. It currently	
		* consists of 7 contigs. The true order of the pieces	
		* is not known and their order in this sequence record is	
		* arbitrary. Gaps between the contigs are represented as	
		* runs of N, but the exact sizes of the gaps are unknown.	
		* This record will be updated with the finished sequence	
		* as soon as it is available and the accession number will	
		* be preserved.	
		* 1 1582: contig of 1582 bp in length	
		* 1583 1582: gap of unknown length	
		* 1683 10849: contig of 9167 bp in length	
		* 10850 10849: gap of unknown length	
		* 10950 26324: contig of 15375 bp in length	
		* 26325 26424: gap of unknown length	
		* 26425 49772: contig of 23348 bp in length	
		* 49773 49872: gap of unknown length	
		* 49873 78964: contig of 29092 bp in length	
		* 78965 79064: gap of unknown length	
		* 79065 107990: contig of 28926 bp in length	
		* 107991 108090: gap of unknown length	
		* 108091 197992: contig of 89902 bp in length.	
FEATURES		Location/Qualifiers	
source		1..197992	
		/organism="Homo sapiens"	
		/mol_type="genomic DNA"	
		/db_xref="taxon:9606"	
		/chromosome="19"	
		/clone="RP11-446K10"	
		1..1582	
misc_feature		/note="assembly_name:Contig4"	
		1683..10849	
misc_feature		/note="assembly_name:Contig6"	

MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McSwan, P., McKernan, K., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Niver, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Plunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Triglio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zemek, L., Zimmer, A. and Zody, N.

Direct Submission
Submitted (32-MAR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 19, 2002 this sequence version replaced gi:17647019.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIGR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L9215
Center clone name: 635_N_19

FEATURES	source	Location/Qualifiers
repeat_region		1..173252 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="18" /map="18" /clone="RP11-635N19" /clone_lib="RPC1-11 Human Male BAC"
repeat_region		2..295 /rpt_family="MER65A-int" 296..330 /rpt_family="(TTTA)n"
repeat_region		331..475 /rpt_family="MER65A-int" complement(553..904) /rpt_family="THE1B"
repeat_region		1274..1618 /rpt_family="MER65A-int"
repeat_region		1619..1955 /rpt_family="MER7A"
repeat_region		1956..2254 /rpt_family="MER65A-int"
repeat_region		2255..2403 /rpt_family="FRAM"
repeat_region		2404..3470 /rpt_family="MER65A-int"
repeat_region		3557..3674 /rpt_family="MER65A-int"
repeat_region		3675..3951 /rpt_family="AluX"
repeat_region		3992..4132 /rpt_family="MER65A-int"
repeat_region		4133..4439 /rpt_family="AluX"
repeat_region		4440..4676 /rpt_family="MER65A-int"
repeat_region		4775..4826 /rpt_family="MER65A-int"
repeat_region		4909..5110 /rpt_family="MER65A-int"
repeat_region		5195..5355 /rpt_family="MER65A-int"
repeat_region		6557..7121 /rpt_family="MER65A-int"


```
repeat_region complement(10651..10682)
repeat_region /rpt family="MER33"
repeat_region 10683..10986
repeat_region /rpt family="AluY"
repeat_region complement(10987..11250)
repeat_region /rpt family="MER33"
repeat_region 11308..11375
repeat_region /rpt family="L2"
repeat_region 12718..12851
repeat_region /rpt family="MIR"
repeat_region complement(12999..13079)
repeat_region /rpt family="L2"
repeat_region 13088..13738
repeat_region /rpt family="L2"
repeat_region 13441
repeat_region /note="Probably a C"
repeat_region complement(14105..14154)
repeat_region /rpt family="L2"
repeat_region 14160..14189
repeat_region /rpt family="(A)n"
repeat_region 14218..14246
repeat_region /rpt family="(TG)n"
repeat_region 14767..15154
repeat_region /rpt family="L1MB3"
repeat_region 15792..16076
repeat_region /rpt family="AluSg"
repeat_region 17327..17519
repeat_region /rpt family="MIR"
repeat_region 17901..18088
repeat_region /rpt family="L2"
repeat_region 18207..18227
repeat_region /rpt family="AT rich"
repeat_region complement(18341..18479)
repeat_region /rpt family="L1PA13"
repeat_region 19251..19304
repeat_region /rpt family="MIR"
repeat_region 19541..19867
repeat_region /rpt family="AluJo"
repeat_region 20031..20144
repeat_region /rpt family="L2"
repeat_region 20197..20415
repeat_region /rpt family="L2"
repeat_region complement(20507..20711)
repeat_region /rpt family="MER30"
repeat_region 22505..22536
repeat_region /rpt family="(CA)n"
repeat_region complement(22609..22669)
repeat_region /rpt family="L2"
repeat_region 23043..23297
repeat_region /rpt family="L2"
repeat_region 23470..23598
repeat_region /rpt family="MER2"
repeat_region complement(23600..23915)
repeat_region /rpt family="AluSx"
repeat_region 23924..23987
repeat_region /rpt family="MER2"
repeat_region 23935..23941
repeat_region /note="<30 qual SNGL region"
repeat_region 23974..24230
repeat_region /note="<30 qual SNGL region"
repeat_region 24223
repeat_region /note="Probably a G"
repeat_region 24242..24425
repeat_region /rpt family="L2"
repeat_region 27566..27586
repeat_region /rpt family="AT rich"
repeat_region complement(28883..28973)
repeat_region /rpt family="MER5A"
repeat_region 32167..32197
repeat_region /rpt family="GC rich"
repeat_region 32284..32313
repeat_region /rpt family="GC rich"
repeat_region 33790..33817

/rpt family="GC rich"
33827..33878
/rpt family="(CGG)n"
complement(33975..34047)
/note="<30 qual SNGL region"
34104..34228
/rpt family="GC rich"
34442..34486
/rpt family="(TG)n"
complement(35364..35660)
/rpt family="AluSg"
complement(35766..35887)
/rpt family="FLAM_A"
37108..37419
/rpt family="AluJb"
37637..37946

Query Match 13.7%; Score 338.4; DB 9; Length 188923;
Best Local Similarity 71.7%; Pred. No. 8.3e-62;
Matches 499; Conservative 0; Mismatches 191; Indels 6; Gaps 4;

QY 1508 AAAAATAGGGTACTAGCCAGCGGGGTGGCTCGCGCCTGTATCCAGCACTTGGGGAG 1567
Db 72247 AGAACTAAGTATATTGGCCAGGCATGGTGGCTCAGCCCATATATCAGGACTTTGGGAG 72306
QY 1568 GCGGAGGCGCGGATCACTTGAGGTCAGAACTTTTCAGACCAGCCTGGTCAACATGGGTG 1627
Db 72307 GCCAAGCGAGGAGATCACTGAGGTGAGGATTTTGAGACTAGCCTGACAAACAT-GGTA 72365
QY 1628 AAACGCGGGCTCTACCAAAAT--ATAAAACCTTAGCTGAGTGTGTAGCGCATGACTGT 1685
Db 72366 AAACCTGTCTCTACTTAAATAACAATAAATAATAGCGGGTGTGGTGGCACACACTGT 72425
QY 1686 AATCCAGCAACTCAGGAGGCTGAGGCA-GAATACTGTTGAACCTTGGAGCGGAGGTT 1744
Db 72426 AATCCAGCTACTCAGGAGGCTGAGGAGGAGTTCGTTGAACCGAGAGGTGGAGTT 72485
QY 1745 GCAGTGAGCTGAGATCGTACCACTGCCTCCAGCTTGGCGACAGCAGCACTCTGCCT 1804
Db 72486 GCAGTGAGCAAGATGGCCACCTGCCTCCAGCTTGGGTGACAGAGAGACTCTGTCT 72545
QY 1805 TAAAAATAAATAAATTTTAAAAAATAAGGGGTACTAATATCTACTTAAAGGATG 1864
Db 72546 CAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 72603
QY 1865 AGGGTTAATTAAGTACACATAAGCCTTAGCGAGTGGCTTATGCTGTATCTCAAC 1924
Db 72604 TATGGTAAATTTAAAAAGAAAAAGAGGCGAGGCAAGTGGCTCACTCTATAATACAGC 72663
QY 1925 ACTTTGGGAGTCTGTGGCGGAGGATCACTTGCAGCCAGGAGTTTGGAGACTAGTCTGGGC 1984
Db 72664 ACTTTGGGAGGAGGAGTGTGGCGGAGTGTGGTGTGTTTGTGTTTATACCAAGTGTGGTGTGC 2044
QY 1985 AACAGAGACATGCTCTATAGTGTGTGTTGTTGTTTGTGTTTGTGTTTGTGTTTGTGTTG 2044
Db 72724 AACATGGCGAACCCTCTCTACCAAAATAAATAAATAAATAAATAAATAAATAAATAA 72783
QY 2045 ACCTGAGTCCAGCTTACTAGGAGGCTGAGTGTGGAGAGTGCCTTGAGCCAGGAGGTC 2104
Db 72784 GTCTATAATCTCAGCTTACTAGGAGGCTGAGTGTGGAGAGTGCCTTGAGCCAGGAGGTC 72843
QY 2105 GAGGCTGAGTGCAGCATGTTGTCCACTGCACCTCCAGCCTGGGCAACACAGCAGACCC 2164
Db 72844 GAGGCTGAGTGCAGTGCAGTGCACCTCCAGCCTGGATGATGATGATGATGATGATGATG 72903
QY 2165 TTGTCTCAAAAAACAACAAAAAGCATCTCAATAAG 2200
Db 72904 CTGCCTCAAAAAACAACAAAAAGGAGGAG 72939

Search completed: August 12, 2005, 02:41:45
Job time : 10570 secs
```